REMARKS

Claims 12-13, 21-24, and 26-32 are pending. Claim 25 has been cancelled, as the amendment to claim 24 makes claim 25 redundant. Support for amended claims 12 and 21-24 and new claims 26-32 derives from the specification and claims as originally filed. For example, computational methods for the generation of primary libraries are described at page 7, line 22, through page 8 line 12, and page 10, line 9 through page 15, line 14. Methods for the generation of secondary libraries from primary libraries are described at page 26, line 27, through page 30, line 27. Support for the synthesis of variant proteins, beginning with the corresponding oligonucleotide sequences using multiple PCR can be found at pages 31-32. Methods for isolating, purifying, and expressing the oligonucleotide sequences as proteins are well known in the art, and are described at pages 41-47 and in the Examples. Accordingly, the amendments do not present new matter and entry is proper.

Rejections under 35 U.S.C. § 112, first paragraph

Claims 12-13 and 21-25 are rejected under 35 U.S.C. § 112, first paragraph for failing to comply with the written description requirement. In rejecting claim 12-13 and 21-25, the Examiner's position appears to be that: (1) the specification does not describe how a library of primary sequences can be received, and (2) that the specification is enabling only for the design of enzymes (*see*, page 7 of the final office action).

In response to the Examiner's first point, Applicants have amended claim 12 to clarify that the primary library of primary variant sequences is provided via a computational processing method based on force field calculations using the coordinates of a target protein. Applicants respectfully submit that this amendment overcomes the Examiner's first point.

In response to the Examiner's second point, Applicants respectfully submit that the specification enables a method for computationally generating a genus of secondary libraries comprising variant sequences in which the starting protein structure (*i.e.* scaffold) can be any protein for which a three dimensional structure is known or can be generated. In addressing the written description requirement under 35 U.S.C. § 112, the Federal Circuit in University of *California v. Eli Lilly and Co.*, 43 USPQ2d 1398, 1406 (Fed. Cir. 1997), stated:

> A description of a genus of cDNAs may be achieved by means of a recitation of a representative number of cDNAs, defined by nucleotide sequence, falling with the scope of the genus or of a recitation of structural features common to the members of the

> genus, which features constitute a substantial portion of the genus. This is analogous to enablement of a genus under Section 112, para. 1, by showing the enablement of a representative number of species within the genus. *See Angstadt*, 537 F.2d at 502-03 (deciding that applicants "are *not* required to disclose *every* species encompassed by their claims their claims even in an unpredictable art and that the disclosure of forty working examples sufficiently described subject matter of claims directed to a generic process) . . . See also *In re Grimme*, 274 F.2d, 949, 952 ("[I]t has been consistently held that the naming of one member of such a group is not, in itself, a proper basis for a claims to the entire group. However, it may not be necessary to enumerate a plurality of species if a genus is sufficiently identified in an application by other appropriate language.").

Applicants respectfully submit that the specification provides written description support and enables a method for computationally generating a genus of secondary libraries comprising variant sequences in which the starting protein structure (*i.e.* scaffold) can be any protein for which a three dimensional structure is known or can be generated. Proteins suitable as starting structures for the generation of secondary libraries are outlined on page 9, line 8 through page 10, line 8.

Regarding the issue that "any scaffold protein" is too broad, it is clear from the description provided in the specification and well known in the art, that a scaffold protein is one that must already have a structure known (e.g. 3D), and is therefore a required input to the method and not a required part of the method. Said scaffold protein may be obtained using well-known structural biology techniques such as X-ray crystallography and NMR spectroscopy, which are commonly made available publicly in the protein databank. Alternatively, a homology model may be generated using one or more structures that are publicly available and algorithms well known in the art. The current invention does not require the user to "generate a scaffold for a single, let alone, for a library of any organisms", as is stated. Rather, the current invention provides methods of protein design, which can be applied to an existing scaffold protein and/or set of corrdinates. In addition, the removal of water, SO_2, and other extraneous non-protein artifacts from an available structure is common practice in the field, and is carried out simply by editing the structure file. Such editing is neither a restriction for the application of the current method nor does it represent undue experimentation.

The computational methods that can be used in generating the secondary libraries are known to those of skill in the art. Finally, two working examples for the generation of variant libraries are provided, using a combination of the computational and experimental methods is described in the specification as filed.

In support of the position that the present invention is enabled, Applicants enclose herewith a number of publications that are both prior and subsequent to the filing date of the present application, that address the enablement of the invention. These are not offered to augment the disclosure of the application; rather, the work is presented to show that present invention is enabled for any protein for which a defined set of coordinates can be generated. *See In re Wilson*, 135 USPQ 442, 444 (CCPA 1962); *Ex parte Obukowicz*, 27 USPQ 2d 1063 (BPAI 1993); *Gould v. Quigg*, 3 USPQ 2d 1302,1305 (Fed. Cir. 1987):

> "it is true that a later dated publication cannot supplement an insufficient disclosure in a prior dated application to render it enabling. In this case the later dated publication was not offered as evidence for this purpose. Rather, it was offered . . . as evidence that the disclosed device would have been operative" printed publications.

In the article "Proteins from Scratch" (DeGrado, Science 278:80-81, 1997, a copy of which is enclosed as Exhibit A), biochemistry professor William F. DeGrado of the University of Pennsylvania School of Medicine, a world-renowned expert in protein structure, folding and design, comments on the computational platform designed by Dahiyat and Mayo in Science 278:82-87 (1997). This platform is an earlier version of the computational platform that has evolved and is claimed herein. Dr. DeGrado states:

> Not long ago, it seemed inconceivable that proteins could be designed from scratch. Because each protein sequence has an astronomical number of potential confirmations, it appears that only an experimentalist with the evolutionary life span of Mother Nature could design a sequence capable of folding into a single, well-defined three dimensional structure. But now on page 82 of this issue, Dahiyat and Mayo describe a new approach that makes de novo protein design as easy as running a computer.

Dr. DeGrado further states (col 1, paragraph 3):

> Thus, the problem of de novo protein design reduced to two steps: selecting a desired tertiary structure and finding a sequence that would stabilize this fold. Dahiyat and Mayo

> have now mastered the second step with spectacular success. They have distilled the rules, insights and paradigms gleaned from two decades of experiments into a single computational algorithm...Thus the rules of ...computational methods for de novo design may now be sufficiently defined to allow the engineering of a variety of proteins.

Holmes (New Scientist, 11 October 1997, enclosed herein as Exhibit B), quoting Dr. Wells, a protein engineer at Genentech, states, "This [the Dahiyat and Mayo work] will stand as a landmark piece of work".

In an article by Borman (Chemical and Engineering Newsletter, October 6, 1997, enclosed as Exhibit C) George D. Rose, formerly professor of biophysics and biophysical chemistry at the Johns Hopkins University refers to the work of Dahiyat and Mayo stating, "Dahiyat and Mayo have taken protein engineering to a new high".

On page 6, the Examiner states that "[a]s a skilled in the art appreciates, to date there are too numerous obstacles for the design of even a single secondary structure of a protein, let along, all or any kinds of proteins".

Professor DeGrado, Dr. Wells and Professor Rose all highlight the concept of the Dahiyat and Mayo computational protein design as a significant breakthrough in science. In addition, since the computational platform uses scaffold proteins (by definition protein structures), there is no discrimination as to the particular "type" of protein, be it enzyme or otherwise. As the Examiner will appreciate, despite the fact that proteins display an enormous array of functions, all are naturally composed of the same 20 amino acids, all governed by the same physico-chemical principles.

Further, Applicants' have designed many proteins that are not "enzymes". For example, see the articles enclosed described computationally designed GCSF (US 6627186 and Luo P et al., Protein Science 11, 1218-1226 (2002); enclosed herein as Exhibits D and E), Interferon Beta (US 6514729, enclosed herein as Exhibit F) and TNF-alpha (US publication No. 2003/138401 and Steed PM et al, Science 301, 1895-1898 (2003); enclosed herein as Exhibits G and H), for example. These non-enzymatic proteins have a variety of structures and have all been successfully designed. Thus, it is improper to limit the scope of this invention to just "enzymes".

The articles, patents and patent applications discussed above, and in particular, the commentary of Professor DeGrado, support the enablement of the methods disclosed in the pending claims. Importantly, the methods apply to proteins in general, regardless of whether the protein is an enzyme, as described in the example, or an antibody, cell surface receptor, or other protein of interest.

Accordingly, Applicants respectfully submit that the specification fully enables the present claims, and respectfully request withdrawal of the rejection under 35 U.S.C. § 112, first paragraph.

Rejections under 35 U.S.C. § 112, second paragraph

Claims 12-13 and 21-25 are rejected under 35 U.S.C. § 112, second paragraph, as being indefinite. In rejecting claim 12-13 and 21-25, the Examiner refers back to the Office Action mailed October 2, 2002. In reviewing the Office Action mailed October 2, 2002, Examiner's first point under paragraph A appears to be that as written, independent claim 12 omits essential steps.

As discussed above, Claim 12 has been amended and thus, the rejection should be withdrawn.

The Examiner other point under paragraph A appears to be that applicants have acquiesced to the rejection regarding the term "probability" by not responding thereto. Applicants respectfully disagree. The term probability is used in connection with a "probability distribution table", which is a term of art. The generation and information comprising the probability tables used in the methods disclosed in claims 12-13 and 21-25 are described throughout the specification, see, e.g., page 23, line 19 through page 30, line 6, and in the Examples. A probability distribution table was generated in the examples, first by analyzing the primary library in the Monte Carlo analysis (Table 3) which generated the probability of each possible amino acid at each position and put it in a table (Table 3), e.g. probability table. Table 4 shows how this distribution can be rounded. Accordingly, Applicants respectfully submit that the term "probability" when used in connection with "probability distribution table" does not connote uncertainty.

Accordingly, Applicants respectfully request withdrawal of the rejection under 35 U.S.C. § 112, second paragraph as set forth in paragraph A.

Applicants acknowledge the Examiner's statement that the rejections under paragraphs B-E are either withdrawn or are moot.

Under paragraph F, the Examiner's first point appears to be that claims 21-25 do not further limit independent claim 12 because claim 12 recites protein variants and claims 21-23 relate to the synthesis of oligonucleotides. Claim 21 has been amended to clarify that the relation between claim 12 and claims 21-23.

The Examiner also rejects the use of the terms "relative amounts", "equimolar amounts" and "correspond". Claims 21-24 have been amended to clarify the metes and bounds of these amounts. Claim 25 has been cancelled. Accordingly, Applicants respectfully request withdrawal of the rejection under 35 U.S.C. § 112, second paragraph as set forth in paragraph F.

Rejections under 35 U.S.C. § 103(a)

Claims 12-13 and 21-25 are rejected under 35 U.S.C. § 103(a) as being unpatentable over Mayo *et al.*, WO/98/47089. This rejection is reiterated for the reasons set forth in the Office Action mailed October 2, 2003. The Examiner's position appears to be that as the multiple PCR technique is known in the art, it would have been obvious to one of skill in the art to synthesize the sequences of Mayo using this technology. Applicants respectfully disagree.

Independent claim 12 discloses a computational method for making secondary libraries from primary libraries comprising the steps a) providing a primary library comprising a plurality of primary variant sequences computationally generated using a force field calculation; b) computationally generating a probability distribution table of amino acid residues in a plurality of variant positions from said primary variant sequences; and, c) computationally combining a plurality of said amino acid residues to generate a secondary library of secondary variant sequences; wherein at least one of said secondary variant sequences is different from said primary variant sequences.

Applicants respectfully submit that Mayo does not disclose methods for computationally generating a secondary library from a primary library. Mayo does, as the Examiner points out, describe the use of rotamer libraries and force field calculations to generate sequences. However, these sequences are the equivalent of the primary sequence

libraries described in the current invention (e.g., Step (A) of claims 26-33). The rotamer library does not constitute such a primary library as described. The rotamers of the Mayo invention and the current invention are defined before any calculation is done, and in fact they are the starting point for an analysis using force field calculations.

To establish a *prima facie* case of obviousness the prior art reference (or references when combined) must teach or suggest all the claim limitations. The teaching or suggestion to make the claimed combination and the reasonable expectation of success must both be found in the prior art, and not based on applicant's disclosure. *In re Vaeck*, 947 F.2d 488, 20 USPQ2d 1438 (Fed. Cir. 1991) M.P.E.P. §2143.

As argued above, Applicants respectfully submit that Mayo does not disclose methods for computationally generating secondary libraries from primary libraries. Accordingly, Applicants respectfully request the rejection under 35 U.S.C. § 103(a) be withdrawn.

The Examiner is invited to contact the undersigned at (415) 781-1989 if any issues may be resolved in that manner.

Respectfully submitted,
DORSEY & WHITNEY LLP

Dated: 5/26/04

By: Renee M Kosslak

Four Embarcadero Center
Suite 3400
San Francisco, California 94111-4187
Telephone: (415) 781-1989
Fax No. (415) 398-3249

Renee M. Kosslak, Reg. No. 47,717
for Robin M. Silva, Reg. No. 38,304
Filed under 37 C.F.R. § 1.34(a)

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

Proteins from Scratch

William F. DeGrado

Not long ago, it seemed inconceivable that proteins could be designed from scratch. Because each protein sequence has an astronomical number of potential conformations, it appeared that only an experimentalist with the evolutionary life span of Mother Nature could design a sequence capable of folding into a single, well-defined three-dimensional structure. But now, on page 82 of this issue, Dahiyat and Mayo (*1*) describe a new approach that makes de novo protein design as easy as running a computer program. Well almost. . .

The intellectual roots of this new work go back to the early 1980s when protein engineers first thought about designing proteins (*2*). At that point, the prediction of a protein's three-dimensional structure from its sequence alone seemed a difficult proposition. However, they opined that the inverse problem—designing an amino acid sequence capable of assuming a desired three-dimensional structure—would be a more tractable problem, because one could "over-engineer" the system to favor the desired folding pattern. Thus, the problem of de novo protein design reduced to two steps: selecting a desired tertiary structure and finding a sequence that would stabilize this fold. Dahiyat and Mayo have now mastered the second step with spectacular success. They have distilled the rules, insights, and paradigms gleaned from two decades of experiments (*3*) into a single computational algorithm that predicts an optimal sequence for a given fold. Further, when put to the test the algorithm actually predicted a sequence that folded into the desired three-dimensional structure. Thus, the rules of protein folding and computational methods for de novo design may now be sufficiently defined to allow the engineering of a variety of proteins.

Dahiyat and Mayo's program divides the interactions that stabilize protein structures into three categories: interactions of side chains that are exposed to solvent, of side chains buried in the protein interior, and of parts of the protein that occupy more interfacial positions. Exposed residues contribute to stability, primarily through conformational preferences and weakly attractive, solvent-exposed polar interactions (*4*). The burial of hydrophobic residues in the well-packed interior of a protein provides an even more powerful driving force for folding. The side chains in the interior of a protein adopt unique conformations, the prediction of which is a large combinatorial problem.

Better than the real thing. The natural zinc finger protein Zif268 (**left**) is stabilized in part by a core of hydrophobic (green) side chains and metal-chelating side chains (red). In the designed protein FSD-1 (**right**), the Zif268 core is retained but the metal-chelating His residues and one of the Cys residues of Zif268 are converted to hydrophobic Phe and Ala residues, thereby extending the hydrophobic core. The fourth metal ligand Cys^{8} is converted to a Lys residue. The apolar portion of this interfacial residue shields the hydrophobic core, whereas its ammonium group is exposed to solvent. The helix is also stabilized by an N-capping interaction (*19*), which presumably also stabilizes the structure.

One important simplifying assumption arose from the early work of Jainin *et al.* (*5*), who showed that each individual side chain can adopt a limited number of low-energy conformations (named rotamers), reducing the number of probable conformers available to a protein. This work was subsequently extended to the design of proteins containing only the most favorable rotamers (*6*). Although the side chains in natural proteins deviate from ideality in a few cases (complicating the prediction of the structures of natural proteins), these deviations need not be considered in the design of idealized proteins. Thus, various algorithms have been developed to examine all possible hydrophobic residues in all possible rotameric states, to find combinations that efficiently fill the interior of a protein. A complementary approach uses genetic methods to exhaustively search for sequences capable of filling a protein core (*7*), and this work has been adapted for the de novo design of proteins (*8*).

Interfacial residues are also quite important for protein stability (*9, 10*). They are often amphiphilic (for example, Lys, Arg, and Tyr) and their apolar atoms can cap the hydrophobic core, while their polar groups engage in electrostatic and hydrogen-bonded interactions.

Until recently, protein designers have frequently concentrated on quantifying the energetics associated with just one of these three types of interactions (*3*). However, de novo design is best approached by simultaneously considering all of the side chains in the protein—unfortunately, a very high-order combinatorial problem. For instance, the volume available to the interior side chains depends on the nature and conformation of the residues at the interfacial positions and vice versa. Dahiyat and Mayo assumed that each of these three features had been adequately quantitated to provide a useful empirical energy function for protein design. Their program combines a number of feaures taken from earlier potential functions and includes a penalty for exposing hydrophobic groups to solvent. Another essential innovation included in their program is an implementation of the Dead-End Elimination theorem, to efficiently search through sequence and side chain rotamer space.

Dahiyat and Mayo's target fold is a zinc finger, a motif with a well-established history in protein structure prediction and design. In an early, prescient paper, Berg correctly inferred that this $His^{2}Cys^{2}$ Zn-binding motif must feature a β-β-α fold that would position the ligating groups in a tetrahedral array around the bound Zn(II) (*11*). Favorable metal ion–ligand interactions together with a small apolar core help stabilize the three-dimensional structure of this compact fold. More recently, Imperiali and co-workers have designed a peptide that folded into this motif, even in the absence of metal ions (*12*). The design included a D-amino acid to stabilize a type II′ turn, and a large, rigid tricyclic side chain that may help consolidate the hydrophobic core. This work was particularly ex-

An enhanced version of this Perspective with links to additional resources is available for Science Online subscribers at www.sciencemag.org

The author is in the Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6059, USA. E-mail: wdegrado@mail.med.upenn.edu

Exhibit A

citing because, before their studies, it was not expected that sequences as short as 25 residues in length could fold into stable tertiary structures.

Now, Dahiyat and Mayo take these studies one step further through the design of a sequence composed of only natural amino acids that adopts the zinc finger motif. As input to their program, they introduced the coordinates of the backbone atoms from the crystal structure of the second domain of the zinc finger protein Zif268. The program then evaluated a total of 10^{62} possible side chain–rotamer combinations to find a sequence capable of stabilizing this fold without a bound metal ion. The resulting protein sequence shares a small hydrophobic core with its predecessor from Zif268. However, in the newly designed protein FSD-1 the core is enlarged through the addition of hydrophobic residues that fill the space vacated by the removal of the metal-binding site (see the figure). This increase in the size of the hydrophobic core together with the enhancements in the propensity for forming the appropriate secondary structure provide an adequate driving force for folding. The designed miniprotein actually folds into the desired structure as assessed by nuclear magnetic resonance spectroscopy, and the observed structure closely resembles the three-dimensional structure of Zif268.

Because of its small size, the protein is marginally stable. A Van't Hoff analysis of the thermal unfolding curve gives a change in the enthalpy (ΔH_{vH}) of approximately −10 kcal/mol, and indicates that the protein is about 90 to 95% folded at low temperatures (*13*). The small value ΔH_{vH} and the lack of strong cooperativity in the unfolding transition are expected for a native-like protein of this very small size (*14*). Thus, FSD-1 is the smallest protein known to be capable of folding into a unique structure without the thermodynamic assistance of disulfides, metal ions, or other subunits. This important accomplishment illustrates the impressive ability of Dahiyat and Mayo's program to design highly optimized sequences.

This new achievement caps a banner year for de novo protein design. Earlier, Regan (*15*) answered the challenge of changing a protein's tertiary structure by altering no more than 50% of its sequence. And although Dahiyat and Mayo have demonstrated that the stabilizing metal-binding site is not necessary in their system, Caradonna, Hellinga, and co-workers (*16*) have made impressive progress in automating the introduction of functional metal-binding sites into the three-dimensional structures of natural proteins. Further, other workers (*17*) have used less automated approaches to successfully introduce functionally and spectroscopically interesting metal-binding sites into de novo designed proteins.

To date, the most computationally intensive protein design problems have been the redesign of natural proteins of known three-dimensional structure. But the new automated approaches open the door to the de novo design of structures with entirely novel backbone conformations. It will be interesting to see if Dahiyat and Mayo's approach of designing an optimal sequence for a given fold is sufficient, or if it will be necessary also to destabilize alternate possible folds. Indeed, when using an earlier version of their algorithm to repack the interior of the coiled coil from GCN4, they had to retain the identity of a buried Asn residue from the wild-type protein. Although the inclusion of this Asn actually destabilized the desired fold, it was nevertheless essential to avoid the formation of alternate, unwanted conformers (*18*). The ability to ask such focused questions will reveal much about how natural proteins adopt their folded conformations while simultaneously allowing the design of entirely new polymers for applications ranging from catalysis to pharmaceuticals.

References and Notes

1. B. I. Dahiyat and S. L. Mayo, *Science* **278**, 82.
2. K. E. Drexler, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5275 (1981); C. Pabo, *Nature* **301**, 200 (1983).
3. W. F. DeGrado, Z. R. Wasserman, J. D. Lear, *Science* **243**, 622 (1989); J. W. Bryson *et al.*, *ibid.* **270**, 935 (1995); M. H. J. Cordes, A. R. Davidson, R. T. Sauer, *Curr. Opin. Struct. Biol.* **6**, 3 (1996).
4. R. Munoz and L. Serrano, *Proteins* **20**, 301 (1994); C. A. Kim and J. M. Berg, *Nature* **362**, 267 (1993); D. L. Minor and P. S. Kim, *Ibid.* **367**, 660 (1994); C. K. Smith, J. M. Withka, L. Regan, *Biochemistry* **33**, 5510 (1994).
5. J. Janin, S. Wodak, M. Levitt, B. Maigret, *J. Mol. Biol.* **125**, 37 (1978).
6. J. W. Ponder and F. M. Richards, *Ibid.* **193**, 775 (1987); J. R. Desjarlais and T. M. Handel, *Protein Sci.* **4**, 2006 (1995); X. Jing, E. J. Bishop, R. S. Farid, *J. Am. Chem. Soc.* **119**, 838 (1997).
7. J. U. Bowie, J. F. Reidhaar-Olson, W. A. Lim, R. T. Sauer, *Science* **247**, 1306 (1990).
8. S. Kamtekar, J. M. Schiffer, H. Xiong, J. M. Babik, M. H. Hecht, *ibid.* **262**, 1680 (1993).
9. K. J. Lumb and P. S. Kim, *ibid.* **271**, 1137 (1996); Y. Yu, O. D. Monera, R. S. Hodges, P. L. Privalov, *J. Mol. Biol.* **255**, 367, (1996).
10. A. C. Braisted and J. A. Wells, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5688 (1996).
11. J. M. Berg, *Ibid.* **85**, 99 (1988).
12. M. D. Struthers, R. P. Cheng, B. Imperiali, *Science* **271**, 342 (1996).
13. This Van't Hoff analysis of the protein is approximate because of the lack of definition of the pre- and posttransition baselines.
14. P. Alexander, S. Fahnestock, T. Lee, J. Orban, P. Bryn, *Biochemistry* **31**, 3597 (1992).
15. S. Dalal, S. Balasubramanian, L. Regan, *Nat. Struct. Biol.* **4**, 548 (1997).
16. A Pinto, H. W. Hellinga, J. P. Caradonna, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5562 (1997); C. Coldren, H. W. Hellinga, J. P. Caradonna, *Ibid.*, p. 6635.
17. B. R. Gibney, S. E. Mulholland, F. Rabanal, P. L. Dutton, *ibid.* **93**, 15041 (1996); M. P. Scott, J. Biggins, *Protein Sci.* **6**, 340 (1997); P. A. Arnold, W. R. Shelton, D. R. Benson, *J. Am. Chem. Soc.* **119**, 3181 (1997); G. R. Dieckman *et al.*, *Ibid.*, p. 6195.
18. P. B. Harbury, T. Zhang, P. S. Kim, T. Alber, *Science* **262**, 1401 (1993); K. J. Lumb and P. S. Kim, *Biochemistry* **34**, 8642 (1995).
19. L. G. Presta and G. D. Rose, *Science* **240**, 1632 (1988); J. S. Richardson and D. C. Richardson, *Ibid.*, p. 1648.

The bigger the better

Size is all-important for cells that want to get around

TINY, free-floating bacteria that may have been among the earliest life forms on Earth were probably not equipped to move about. So says an American biophysicist who has calculated that microbes below a certain size cannot derive any benefit from being able to steer themselves around. The work adds a constraint to evolution that should apply anywhere in the Universe.

Exactly how small a cell can be is controversial. Although most biologists believe no bacterial cells less than 0·2 micrometres long exist, some claim to have evidence for cells a tenth of this size. The debate hotted up last year, after NASA scientists claimed to have found such miniature microbes fossilised inside a Martian meteorite (This Week, 17 August 1996, p 4).

David Dusenbery of the Georgia Institute of Technology in Atlanta says he has come across part of the answer by studying the physics of the microbial environment, rather than the bacteria themselves. "How much information an organism can gather about its environment, and how it then acts, clearly has physical constraints," he says.

Dusenbery found that on the nanometre scale, the ability of a bacterium to swim in a particular direction can be fairly accurately modelled with just a handful of equations. These define how the organisms would react in different conditions of light, temperature and chemistry. It turns out that in almost every circumstance, free-floating cells below 0·6 micrometres simply do not benefit from swimming, no matter how powerful their "motors".

Nice mover: large bacteria have in-built "motors"

Dr Kari Lounatmaa/Science Photo Library

The reason, says Dusenbery, is that such organisms live in a physical world that defies our everyday intuition. For example, on small scales, water is a very viscous fluid. If a very small bacterium stopped swimming even momentarily, it would grind to a halt in a few atoms' length. Furthermore, the random buffeting by water molecules would be more than enough to push the cells off-course.

Dusenbery concludes that if motility can't do a small cell any good, it should not evolve in such cells. To test that prediction, he compiled a list of 218 genera of free-floating bacteria, about half motile and half non-motile. In general, motile bacteria were larger than their less active kin, and none of the motile bacteria fell below the 0·6 micrometre cut-off (*Proceedings of the National Academy of Science*, vol 94, p 10949). Dusenbery is confident that the same will be true on any other planet: "As long as the laws of physics were the same, the same rules apply."

"It's intriguing work—he seems to have gone into all the aspects of environment you can imagine," comments Robert Macnab of Yale University, who studies the biology of bacterial motion. But he adds that the protein motors that power modern bacteria are incredibly complex, requiring more than 60 genes. That alone might explain why only larger, more complex bacteria are capable of independent motion. **Philip Cohen**

First-ever designer protein fits like a glove

A MOLECULAR tailor's shop has turned out the world's first made-to-order protein. The feat opens the door to a future in which scientists might design bespoke proteins for drugs or industrial catalysts.

A protein is a long chain of amino acid building blocks, tightly folded on itself like a scrunched-up piece of string. Electrically charged parts remain exposed to surrounding water molecules, while neutral parts are tucked away inside. With interactions between many thousands of atoms, proteins are so complex that scientists have not been able to predict their folded shapes from their amino acid sequences.

Now, after five years of studying the complex atomic interactions in proteins, Stephen Mayo of the California Institute of Technology in Pasadena has developed a computer program that selects the sequence of amino acids required to yield a folded protein of a specific shape. With his colleague Bassil Dahiyat, he has used the program to create the first fully tailored protein.

The protein, containing 28 amino acids, could be put together in 10^{27} different ways. But the researchers used the program to predict the sequence that would give a shape that mimics an existing protein type called the "zinc finger". In last week's *Science* (vol 278, p 82), the researchers say that when they synthesised the protein, it fitted the zinc-finger shape almost perfectly—proof that protein design is really possible. "This will stand as a landmark piece of work," says Jim Wells, a protein engineer at Genentech in San Francisco.

Mayo hopes eventually to design proteins up to 150 amino acids long, the size of most natural proteins. This would enable biotechnologists to design heat-stable enzymes for industry and make more effective drugs. "It would be an awesome achievement to design a protein that would do what you want it to," says Mayo. Bob Holmes

Exhibit B

Development of a cytokine analog with enhanced stability using computational ultrahigh throughput screening

PEIZHI LUO, ROBERT J. HAYES, CHERYL CHAN, DIANE M. STARK, MARIAN Y. HWANG, JONATHAN M. JACINTO, PADMAJA JUVVADI, HELEN S. CHUNG, ANIRBAN KUNDU, MARIE L. ARY, AND BASSIL I. DAHIYAT

Xencor, Inc., Monrovia, California 91016, USA

(RECEIVED November 13, 2001; FINAL REVISION February 11, 2002; ACCEPTED February 13, 2002)

Abstract

Granulocyte-colony stimulating factor (G-CSF) is used worldwide to prevent neutropenia caused by high-dose chemotherapy. It has limited stability, strict formulation and storage requirements, and because of poor oral absorption must be administered by injection (typically daily). Thus, there is significant interest in developing analogs with improved pharmacological properties. We used our ultrahigh throughput computational screening method to improve the physicochemical characteristics of G-CSF. Improving these properties can make a molecule more robust, enhance its shelf life, or make it more amenable to alternate delivery systems and formulations. It can also affect clinically important features such as pharmacokinetics. Residues in the buried core were selected for optimization to minimize changes to the surface, thereby maintaining the active site and limiting the designed protein's potential for antigenicity. Using a structure that was homology modeled from bovine G-CSF, core designs of 25–34 residues were completed, corresponding to 10^{21}–10^{28} sequences screened. The optimal sequence from each design was selected for biophysical characterization and experimental testing; each had 10–14 mutations. The designed proteins showed enhanced thermal stabilities of up to 13°C, displayed five- to 10-fold improvements in shelf life, and were biologically active in cell proliferation assays and in a neutropenic mouse model. Pharmacokinetic studies in monkeys showed that subcutaneous injection of the designed analogs results in greater systemic exposure, probably attributable to improved absorption from the subcutaneous compartment. These results show that our computational method can be used to develop improved pharmaceuticals and illustrate its utility as a powerful protein design tool.

Keywords: Protein design; computational screen; stability; cytokines; granulocyte-colony stimulating factor

Many techniques have been used in the design of new and improved proteins. In vitro directed evolution methods such as phage display, DNA shuffling, and error-prone PCR are widely used. Rational design approaches continue to be applied, and strategies that combine both are now being used. Successful designs include enzymes (Chen and Arnold 1991; Stemmer 1994; Zhao et al. 1998) and other proteins (Crameri et al. 1996), as well as therapeutically useful proteins such as hormones and cytokines (Lowman and Wells 1993; Heikoop et al. 1997; Grossmann et al. 1998; Chang et al. 1999). The experimental techniques involve the generation and screening of libraries of random protein sequences. However, the number of sequences that can be screened experimentally is limited (about 10^{14} for library panning and 10^{7} for high throughput screening). Libraries of this size allow for the simultaneous modification of only about 10 residues.

Reprint requests to: Bassil I. Dahiyat, Xencor, Inc., 111 W. Lemon Avenue, Monrovia, California 91016, USA; e-mail: baz@xencor.com; fax: (626) 256-3562.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.4580102.

Exhibit E

Computational methods have also been used that perform in silico screening of protein sequences (Hellinga and Richards 1994; Desjarlais and Handel 1995; Dahiyat and Mayo 1996, 1997a; Street and Mayo 1999; Jiang et al. 2000; Kraemer-Pecore et al. 2001; Pokala and Handel 2001). Exploiting the efficiency and speed of computers, these methods can randomly screen a vast number of sequences (up to 10^{80}), allowing for the simultaneous consideration and modification of more than 60 residues. Searching such large sequence spaces drastically improves the possibility of finding novel protein sequences with improved properties.

Investigators have recently developed a computational screening method that finds the optimal sequence for a defined three-dimensional structure, allowing all or part of the sequence to change (Dahiyat and Mayo 1996). This method, termed Protein Design Automation (PDA), scores the fit of sequences to the three-dimensional structure using physical-chemical potential functions that model the energetic interactions of protein atoms, including steric, solvation, and electrostatic interactions. PDA couples these potential functions with a highly efficient search algorithm to accurately screen up to 10^{80} sequences. Because the screening is performed in silico, multiple simultaneous mutations can be made, and novel sequences that are very different from wild type can be discovered. The method has been validated by numerous experimental tests and has resulted in the design of new proteins with improved stability and conformational specificity, and novel activity (Dahiyat and Mayo 1996, 1997a; Malakauskas and Mayo 1998; Strop and Mayo 1999; Shimaoka et al. 2000; Bolon and Mayo 2001; Marshall and Mayo 2001).

PDA also has the advantage of being able to control the location and type of mutations. For example, the design can be limited to the hydrophobic core. Mutations in the core can produce significant improvements in protein stability but do not change binding epitopes on the surface of the molecule. Thus, the molecular surface can be kept identical to the native structure, retaining biological activity and limiting toxicity and antigenicity. This feature is particularly important in the design of therapeutic proteins.

We wanted to take advantage of these features of PDA and explore its utility in the design of improved pharmaceuticals. We therefore used PDA as an ultrahigh throughput screen for improved analogs of a therapeutic protein, granulocyte-colony stimulating factor (G-CSF). G-CSF is a hematopoietic growth factor of 174 residues that induces differentiation and proliferation of granulocyte-committed progenitor cells. It is used clinically to treat cancer patients and alleviate the neutropenia induced by high-dose chemotherapy. G-CSF belongs to the class of long-chain four-helix bundle cytokines that bind asymmetrically to homodimeric complexes of cell-surface receptors to initiate an intracellular signaling cascade. Their structural similarity allows the design strategy chosen for G-CSF to be immediately applicable to the other four-helix bundle cytokines (human growth hormone, erythropoietin, the interleukins, and interferon-α/β—all clinically important compounds) and thus broadens the potential impact of the results.

Although the cytokines are functionally very efficacious, their pharmacological properties are not ideal. For example, G-CSF, like most proteins, is not absorbed orally to any significant extent and must be administered by frequent (daily) injections throughout the course of treatment. It also has limited stability and strict formulation and storage requirements, including the need to be kept refrigerated. Thus, there is significant interest in developing analogs with improved pharmacological properties.

We sought to use PDA to improve the physicochemical characteristics of G-CSF. Improving these properties can make a molecule more robust, enhance its shelf life, or make it more amenable to use in alternate delivery systems and formulations. It can also affect clinically important features such as pharmacokinetics and result in a drug that is safer for human use. Our design strategy was to optimize the core to improve the stability and solution properties of G-CSF while preserving receptor binding and biological activity.

The template structure used for in silico screening was a homology model of human G-CSF in which the human sequence was mapped onto bovine G-CSF. We designed several novel core sequences, cloned and expressed them, characterized their stabilities, tested them for functional activity both in vitro and in vivo, and studied their pharmacokinetics in monkeys. The designed proteins showed enhanced thermal stabilities, displayed five- to 10-fold improvements in shelf life, and were biologically active both in cell proliferation assays and in a neutropenic mouse model. Subcutaneous injection of the most stable variant in monkeys also resulted in greater systemic exposure, probably attributable to improved absorption from the subcutaneous compartment. These results indicate that PDA has great potential as a powerful in silico tool in the design of improved pharmaceutical proteins.

Results and Discussion

Homology modeling

The crystal structure of bovine G-CSF (PDB record 1bgc) (Lovejoy et al. 1993) was used as the starting point for modeling because the crystal structure of human G-CSF (PDB record 1rhg) (Hill et al. 1993) is at a lower resolution and is missing key fragments, including a structurally important disulfide bond between positions 64 and 74. Bovine G-CSF is a good model for human G-CSF because the sequences are the same length and 142 of 174 amino acids are identical (82%). The residues that differ in the bovine sequence were replaced with the human residues for those

positions, and the conformations of the replaced side chains were optimized using PDA. Most of the replaced residues were solvent exposed, thereby introducing little strain into the structure and allowing typical PDA parameters to be used for conformation optimization. One substitution, however, was at a buried site, G167V, and clashed sterically with a nearby disulfide bond. To accommodate the larger Val, the side-chain conformation at this position was optimized using a less restrictive van der Waals scale factor (0.6 instead of 0.9). The entire structure was then briefly minimized to relax the strain. The final structure that served as the template for all the designs is shown in Figure 1.

Core designs

Unlike many experimental sequence screening methods, PDA allows control over which residues are allowed to change. Core residues were selected because optimization of these positions can improve stability yet minimize changes to the molecular surface, thus limiting the designed protein's potential for antigenicity. Ala scanning studies of G-CSF indicate one or two binding sites on the protein surface that are probably responsible for granulopoietic activity (Reidhaar-Olson et al. 1996; Young et al. 1997) (Fig. 1). Although recent crystallographic studies of G-CSF complexed to its receptor show only one binding site in a novel 2:2 complex (Horan et al. 1996; Aritomi et al. 1999), both sites were avoided in the core designs to ensure preservation of function.

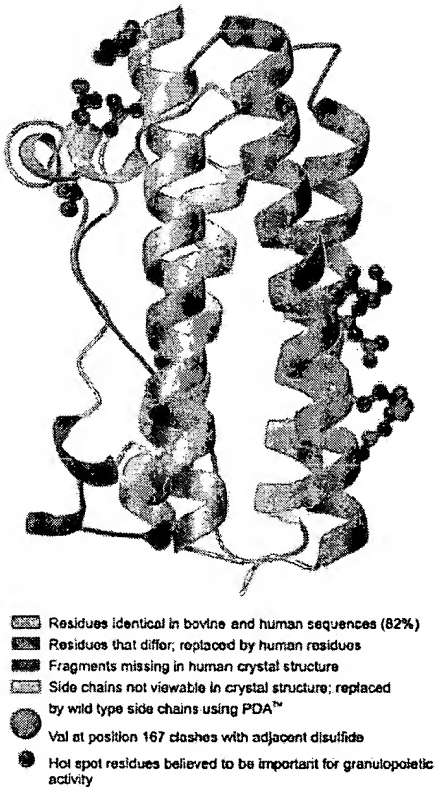


Fig. 1. Template structure of hG-CSF used for Protein Design Automation (PDA) designs. The human sequence was homology modeled onto the bovine crystal structure (PDB record 1bgc). The residues that differ in the bovine sequence or were not present in the bovine crystal structure were replaced with the residues from the human sequence. The conformations of the replaced side chains were optimized using PDA (the larger Val at position 167 was optimized using a less restrictive van der Waals scale factor), and the entire structure was energy minimized for 50 steps.

Two PDA design calculations were run: a deep core design that included residues deeply buried in the interior of the protein and an expanded core design (exp_core) that also included less buried peripheral core residues. The deep core design had 26 core positions that were allowed to vary (shown yellow and gold in Fig. 2), whereas exp_core had 34 (shown yellow and turquoise in Fig. 2). Only hydrophobic amino acids were considered at the variable core positions. These included Ala, Val, Ile, Leu, Phe, Tyr, and Trp. Gly was also allowed for the variable positions that had Gly in the bovine wild-type structure (positions 28, 149, 150, and 167). Met and Pro were not allowed.

Optimal sequences

The optimal sequences selected by PDA are also shown in Figure 2. The optimal sequence from the deep core design had 10 mutations (named core10), and the optimal exp_core sequence had 11 (named exp_core11); thus, 33%–38% of the variable residues changed their identities. Eight of the mutated positions changed to the same amino acid in both designs. Changing the set of design positions can significantly impact the amino acid selected at a given position. For example, in the deep core design, Leu89 retains the same amino-acid identity and conformation as wild type. However, in the exp_core design, when Leu92 is also allowed to vary, both positions (Leu89 and Leu92) mutate to Phe, indicating a coupling between these two core residues. The modeled structure of the sequence selected in the deep core design (core10) is shown in Figure 3.

Native human G-CSF (met hG-CSF) and the optimal sequence from each of the core designs were cloned, expressed in *Escherichia coli*, and purified for experimental studies.

Thermal stability

The far-ultraviolet (UV) circular dichroism (CD) spectra for met hG-CSF and the designed proteins were nearly identical to each other and to published spectra for met hG-CSF (Reidhaar-Olson et al. 1996; Young et al. 1997), indicating highly similar secondary structure and tertiary folds (data

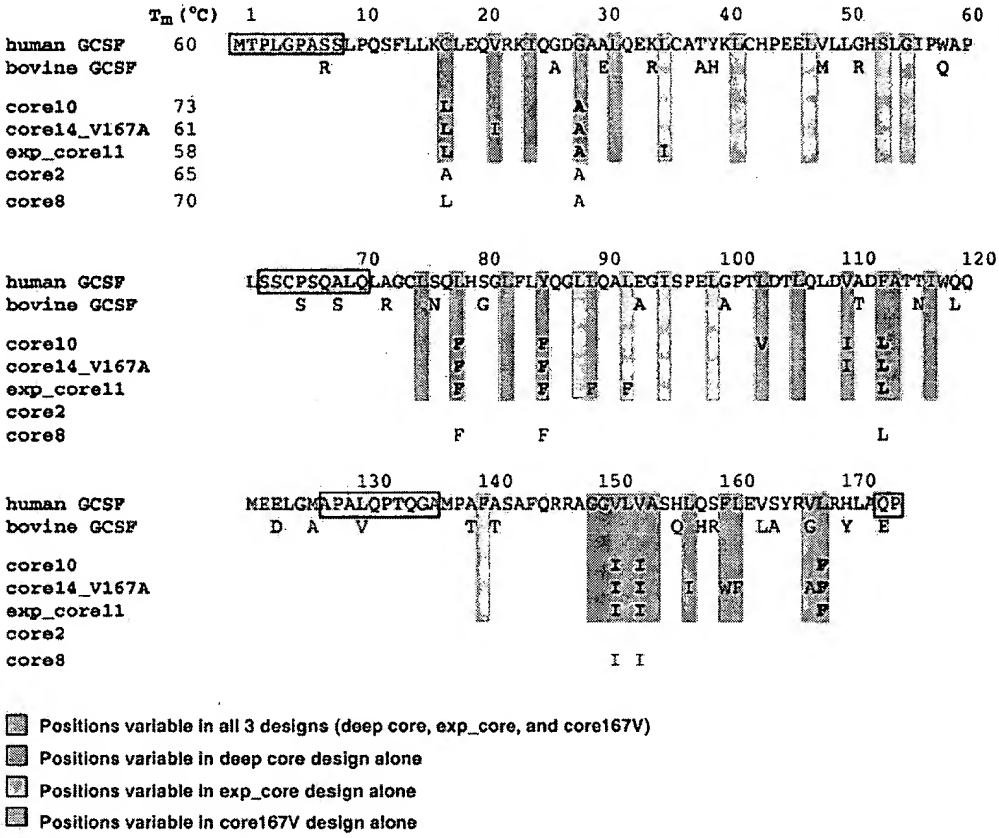


Fig. 2. Sequences of hG-CSF analogs. Native human and bovine sequences are shown at the *top*. The fragments missing in the crystal structure of the human sequence are shown boxed. Variable positions are colored. The deep core design had 26 variable positions, exp_core had 34, and core167V had 25. The optimal sequence from each design is shown. Letters indicate core residues that mutated relative to native hG-CSF; blanks indicate no change. Positions that changed to the same amino acid in all three core designs are indicated in bold. Core2 and core8 sequences were not obtained from PDA calculations but were derived by reverting some of the core10 mutations to wild type. Melting temperatures (T_ms) obtained for the designed proteins are also shown.

not shown). Thermal denaturation was monitored at 222 nm, and the melting temperatures (T_ms) were derived from the derivative curve of the ellipticity at 222 nm versus temperature (Fig. 4). Thermal denaturation of G-CSF and its variants is irreversible; however, T_m can be used to quickly assess the relative stability of different mutants. Stability under storage conditions, which is more relevant clinically, was evaluated with shelf-life studies (see below).

The T_m for met hG-CSF was 60°C, identical to that reported in other studies (Kolvenbach et al. 1997). Core10 showed an increase in stability of 13°C, whereas the T_m of exp_core11 was very similar to wild type (Fig. 2 and Fig. 4). The increased stability seen with core10 may be attributable to improved packing interactions and optimized hydrophobic burial of side chains. Other possibilities include decreased aggregation resulting from elimination of the free cysteine at position 17. The Gly to Ala mutation at position 28 caused a significant improvement in helical propensity that could also be the source of the improved stability.

Identifying critical mutations using derived sequences

To differentiate between these possibilities, two additional sequences derived from the core10 mutant sequence were made and their T_ms measured. One of these (core8) was identical to core10 except that two mutations distant from the others were reverted to wild type (L103V and V110I). These were the two positions that did not mutate in exp_core11. The T_m of core8 was 70°C, similar to core10, indicating that the mutations at 103 and 110 were not responsible for core10's improved stability.

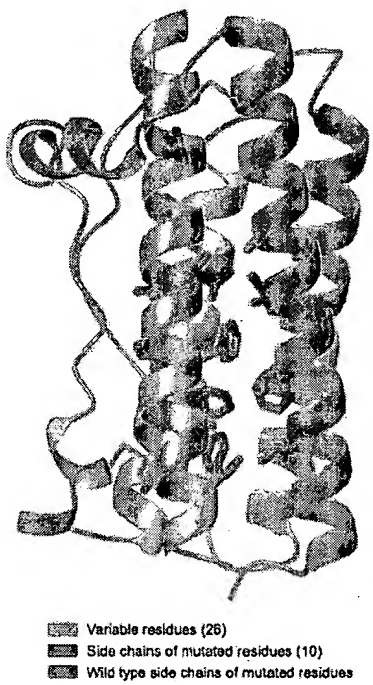


Fig. 3. Modeled structure of hG-CSF analog (core10) obtained from deep core design. Twenty-six core residues were allowed to vary; computational screening with PDA resulted in 10 mutations: C17L, G28A, L78F, Y85F, L103V, V110I, F113L, V151I, V153I, and L168F.

To determine the importance of the other mutations, another sequence was made (core2) that contained only two of the core10 mutations, G28A and C17A; all other residues were identical to wild type (Fig. 2). The T_m' of core2 was 5°C higher than wild type, indicating that improvements in helical propensity and the elimination of a free cysteine are important for heightened thermostability. The remainder of the increase in T_m seen for core10 may be attributable to improved packing interactions and increased hydrophobic burial.

Storage stability

Increased shelf life is important for distribution and storage and is a desirable feature for G-CSF and other protein drugs. Because aggregation and chemical degradation are the predominant mechanisms of inactivation of G-CSF (Herman et al. 1996), shelf life was estimated by incubating the proteins at elevated temperature and then using size-exclusion chromatography to observe the disappearance of monomeric protein. Chemical degradation was estimated using reverse phase chromatography (data not shown). Core2 and core10 showed five- and 10-fold improvements in storage stability, respectively, at 50°C (Fig. 5). Rate constants were determined by a first order exponential fit of the fraction monomer remaining/time curves using KaleidaGraph (Synergy Software).

Biological activity

Granulopoietic activity was determined in vitro by quantitating cell proliferation as a function of protein concentration in murine lymphoid cells transfected with the gene for the human G-CSF receptor. The designed proteins were as

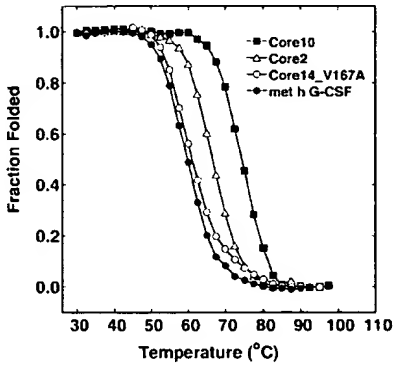


Fig. 4. Thermal stability of hG-CSF analogs. Thermal stability was assessed by monitoring the temperature dependence of the circular dichroism spectral signal at 222 nm. Melting temperatures (T_ms) were derived from the derivative curve of the ellipticity at 222 nm versus temperature. Core10 and core2 showed increases in T_m of 13°C and 5°C, respectively, over native met hG-CSF.

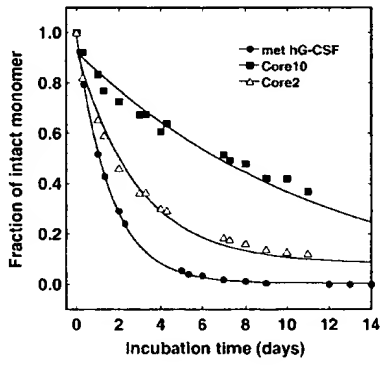


Fig. 5. Shelf life of hG-CSF analogs. Shelf life was estimated by incubating the proteins at elevated temperature (50°C) and using size exclusion chromatography to observe disappearance of monomeric protein. Rate constants were determined by a first order exponential fit of the fraction monomer remaining/time curves. Core2 and core10 showed five- and 10-fold improvements in storage stability, respectively, over met hG-CSF controls.

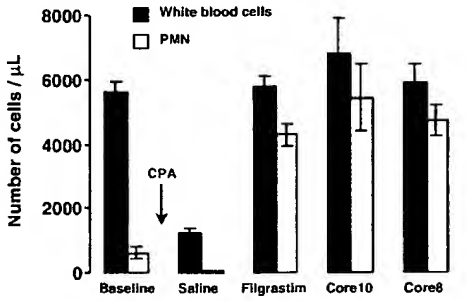


Fig. 6. In vivo granulopoietic activity of hG-CSF analogs. Mice were rendered neutropenic with a single intraperitoneal injection of 200 mg/kg cyclophosphamide (CPA). Beginning 24 h later and for 4 consecutive days, the mice were given a daily intravenous injection of 100 µg/kg of native hG-CSF (filgrastim, Amgen), an hG-CSF analog, or saline. On day 5, granulopoietic activity was determined by counting the number of white blood cells and polymorphonuclear neutrophils (PMN). The designed analogs (core8 and core10) were as effective as controls in eliciting a granulopoietic response.

active as wild-type hG-CSF (data not shown). The designed analogs were also as effective as wild type in increasing white blood cell and polymorphonuclear neutrophil levels in the neutropenic mouse (Fig. 6). Neutropenia, characterized by an abnormally low level of neutrophils in the blood, was induced by injection of cyclophosphamide. Reversal of this effect by the designed analogs shows that granulopoietic activity was also retained in vivo.

Pharmacokinetics

The pharmacokinetics of core10 and native hG-CSF (filgrastim, Amgen) was studied in cynomolgus monkeys after a single subcutaneous or intravenous injection of 5 µg/kg and after daily subcutaneous injections of 5 µg/kg for 28 d. Analysis of the serum concentration-time curves shows that subcutaneous injection of the designed analog results in greater systemic exposure (area under concentration-time curve, AUC) than the same dose of wild-type hG-CSF (Fig. 7B). This was true after a single dose on day 1 (78.8 vs. 54.6 ng-h/mL, data not shown), as well as on day 28 (37.2 vs. 17.4 ng-h/mL). There were no measurable differences in serum half-life. In the intravenous study, however, the half-life of core10 was three-fold shorter (1 vs. 3 h), and the AUC was significantly less (54.7 vs. 117.4 ng-h/mL), indicating that core10 is cleared faster (Fig. 7A). Taken together, these data indicate that the designed analog is absorbed more quickly from the subcutaneous compartment (absorption could not be measured directly given the small number of data points at early times). Improved absorption may be attributable to decreased aggregation or association of the designed protein. The increased monomer lifetime and decreased aggregation seen in our shelf-life studies and the improved thermal stability of the native conformation observed for core10 indicate a decrease in aggregation in the subcutaneous compartment. This possibility is supported by the fact that other protein therapeutics engineered for reduced aggregation also show faster absorption rates. For example, insulin Lispro and other rapid-acting insulin analogs that were designed to decrease their tendency to self-associate are absorbed faster than regular insulin after subcutaneous injection (Howey et al. 1994; Home et al. 1999).

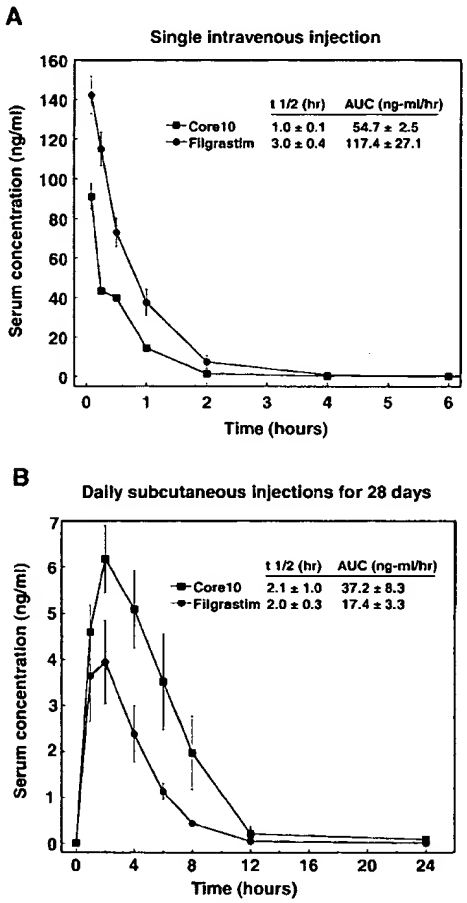


Fig. 7. Pharmacokinetics of hG-CSF analogs. Plasma concentrations of a designed hG-CSF analog or wild-type hG-CSF (filgrastim, Amgen) were determined after administration in cynomolgus monkeys. (*A*) Animals were given a single intravenous injection of 5 µg/kg or (*B*) daily subcutaneous injections of 5 µg/kg for 28 d. Noncompartmental analysis of the serum concentration-time curves shows that subcutaneous injections of the core10 analog resulted in greater systemic exposure (area under concentration-time curve, AUC) than the same dose of wild-type hG-CSF, whereas there was no change in serum half-life ($t_{1/2}$). In the intravenous study, the AUC was significantly less and the $t_{1/2}$ three-fold shorter, indicating that core10 was cleared faster.

Comparison to published G-CSF variants

In vitro and cassette mutagenesis studies have shown that alterations of the N-terminal region of G-CSF can lead to improved granulopoietic activity (Kuga et al. 1989; Okabe et al. 1990). Point mutations at Cys17 have also been found to affect shelf life; replacement with Ala led to an increase, Ser had no effect, and large residues (Ile, Tyr, Arg) led to a decrease (Ishikawa et al. 1992). In contrast, our core10 sequence, which has a large residue (Leu) at this position, showed an improved shelf life. This may be explained by the observation that in a Cys17Leu point mutant, Leu's side chain would clash with the aromatic ring of the nearby Phe at position 113. This steric clash does not occur in core10, however, because the Phe at 113 is replaced by Leu and, in compensation for this change, two nearby Leu's become Phe's (at positions 78 and 168). Thus, multiple mutations allow complementary repacking of the hydrophobic core in the core10 mutant and may be responsible for its enhanced stability and shelf life.

Significant improvements in thermal stability were also observed when the seven helical Gly residues in G-CSF were replaced with Ala to form point, double, and triple mutants (Bishop et al. 2001). Substitutions at positions 26, 28, 149, and 150 were the most effective. The investigators attributed the stabilizing effect to the enhancement in α-helical propensity associated with the Gly/Ala substitutions. These data support our suggestion that the heightened thermal stability seen with our mutants (which also contain a Gly/Ala substitution at position 28) is at least in part attributable to an improvement in helical propensity.

Probing the robustness of PDA with a homology modeled core position

As pointed out previously, the homology modeling of human G-CSF onto the bovine structure was straightforward for the most part because the replaced residues were primarily solvent exposed and no rearrangement of the backbone was necessary. The change at one core position, however, G167V, induced a steric clash and energy minimization of the entire protein was used to relieve the strain. We decided to assess the impact of this manipulation by doing an additional design (core167V) in which the variable residues were essentially the same as in the deep core design except that position 167 was also allowed to vary. We found that Val167 mutated to Ala (the other mutations were essentially the same as for core10). To probe the plasticity of the core, instead of using this PDA optimal sequence, which only had two mutations in this region, we ran experiments on another high-scoring sequence (core14_V167A) that had additional mutations (14 total, including L157I, F160W, and L161F). This sequence was chosen because it balanced an extensive number of mutations with a relatively high design score. Although it ranked 21st in the sequence energy list and was 2 kcal/mole less favorable than the optimal sequence, it was still biologically active and as stable as wild type (T_m of 61°C) (Figs. 2, 4). This indicates that optimization with PDA is fairly robust, and that the protein core can be quite plastic and can accommodate large changes without sacrificing stability or function.

Conclusions

PDA is a powerful ultrahigh throughput computational screening method. Its ability to screen up to 10^{80} sequences and allow multiple simultaneous mutations significantly increases the likelihood of finding new and improved proteins. In this study, PDA was used to develop improved analogs for a therapeutically important protein, hG-CSF. The novel proteins showed enhanced thermal stabilities and shelf life while retaining biological activity. Analysis of the mutants and results obtained with derived sequences indicates that the heightened stability is attributable to improvements in helical propensity and the elimination of a free cysteine; improved core packing and optimized hydrophobic burial of side chains may also be important. Pharmacokinetic studies indicate that subcutaneous injection of the most stable variant results in greater systemic exposure, probably attributable to improved absorption from the subcutaneous compartment.

These results show that PDA can be successfully applied to proteins of therapeutic interest. They also illustrate the value of its precise control over the site and type of mutations, allowing for the rational design of desired properties such as improved stability and pharmacokinetics and the elimination of undesirable ones such as toxicity and antigenicity. These features are particularly important in the design of therapeutic proteins. PDA thus has great potential as a powerful in silico tool for therapeutic protein design.

Materials and methods

Template structure preparation

The template structure for the designed proteins was produced by homology modeling using the crystal structure of bovine G-CSF (Brookhaven Protein Data Bank code 1bgc) as the starting point. The program BIOGRAF (Molecular Simulations Inc., San Diego, CA) was used to generate explicit hydrogens on the structure, which was then minimized for 50 steps using the conjugate gradient method and the Dreiding II force field (Mayo et al. 1990). The residues that differ in the bovine sequence or were not present in the bovine crystal structure were replaced with the human residues for those positions. The conformations of the replaced side chains were optimized using PDA (Dahiyat and Mayo 1997a,b), and the entire structure was minimized again for 50 steps. This minimized structure was used as the template for all the designs.

Protein design

Analogs of hG-CSF were designed by simultaneously optimizing residues in the buried core of the protein using PDA. The computational details, residue classification, potential functions, and parameters used for van der Waals interactions, solvation, and hydrogen bonding are described in previous work (Dahiyat and Mayo 1996, 1997a). An expanded version of the backbone-dependent rotamer library of Dunbrack and Karplus (Dunbrack and Karplus 1993) was used in all the calculations. The global optimum sequence from each design was selected for characterization and experimental testing, except for core167V in which the 21st ranked sequence was used. Calculations were generally performed overnight using 16 processors of an SGI Origin 2000 with 32 R10000 processors running at 195 MHz. The length of the runs varied from 1 to several hours of CPU time.

Cloning and expression

A gene for met hG-CSF was synthesized from partially overlapping oligonucleotides (~100 bases) that were extended and PCR amplified. Codon usage was optimized for *E. coli* and several restriction sites were incorporated to ease future cloning. These partial genes were cloned into a vector and transformed into *E. coli* for sequencing. Several of these gene fragments were then cloned into adjacent positions in an expression vector (pET17 or pET21) to form the full-length gene for met hG-CSF (528 bases) and transformed into *E. coli* for expression. Protein was expressed in *E. coli* in insoluble inclusion bodies and its identity was confirmed by immunoblot of SDS-PAGE using a commercial mAb against hG-CSF.

Refolding, purification, and storage

The protein inclusion bodies were solubilized in detergent and refolded in the presence of $CuSO_4$ to promote formation of native disulfide bonds (Lu et al. 1992). A size-exclusion column (10 mm × 300 mm loaded with Superdex prep 75 resin purchased from Pharmacia) was loaded with protein and eluted at a flow rate of 0.8 mL/min using the column buffer (100 mM Na_2SO_4, 50 mM Tris, pH 7.5). The peaks were monitored at dual wavelengths of 214 nm and 280 nm. Albumin, carbonic anhydrate, cytochrome C, and aprotinin were used to calibrate the molecular size of proteins versus elution time. The monomeric peak that elutes around the expected elution time for each protein was collected and the buffer was exchanged into 10 mM NaOAc at pH 4 for biophysical characterization. For long-term storage, a buffer of 5% sorbitol, 0.004% Tween 80, and 10 mM NaOAc at pH 4 was used. A pH of 4 was chosen for these buffers to be consistent with the commercial formulation of hG-CSF (Amgen), which was used as a control. The proteins were >98% pure as judged by reversed phase high performance liquid chromatography (HPLC) on a C4 column (3.9 mm × 150 mm) with a linear acetonitrile-water gradient containing 0.1% TFE. The identities of all proteins were confirmed by comparing the molecular mass measured by mass spectrometry with corresponding molecular mass calculated using the protein sequences.

Spectroscopic characterization

Protein samples were 50 μM in 50 mM sodium phosphate at pH 5.5. Concentrations were determined using UV spectrophotometry. Protein structure was assessed by CD. CD spectra were measured on an Aviv 202DS spectrometer equipped with a Peltier temperature control unit using a 1-mm path length cell. Thermal stability was assessed by monitoring the temperature dependence of the CD signal at 222 nm (Kolvenbach et al. 1997). A buffer of 10 mM NaOAc was used at pH 4.0 and data were collected every 2.5°C with an averaging time of 5 sec and an equilibration time of 3 min. Thermal denaturation curves were smoothed using KaleidaGraph. The melting temperature (T_m) of each protein was derived from the derivative curve of the ellipticity at 222 nm versus temperature. The T_m values were reproducible to within 2°C for the same protein at the concentrations used.

Storage stability

The storage stability of the designed proteins was assessed by incubation at both 37°C and 50°C under solution conditions identical to that used in the commercial formulation of hG-CSF (filgrastim, Amgen). Because aggregation and chemical degradation are the predominant mechanisms of inactivation of G-CSF (Herman et al. 1996), accelerated degradation was followed by observing the disappearance of monomeric protein with both size-exclusion and reverse-phase chromatography. Rate constants for shelf-life estimation were determined by a first-order exponential fit of the fraction monomer remaining/time curves using KaleidaGraph (Synergy Software).

Cell proliferation assay

Granulopoietic activity was measured by quantifying cell proliferation as a function of protein concentration using Ba/F3 (murine lymphoid) cells stably transfected with the gene encoding the human Class 1 G-CSF receptor (Avalos et al. 1995). Cell proliferation was detected by 5-bromo-2′-deoxyuridine (BrdU) incorporation quantified by a BrdU-specific ELISA kit (Boehringer Mannheim).

In vivo biological activity

Granulopoietic activity was determined in the neutropenic mouse (Hattori et al. 1990). C57BL/6 mice were rendered neutropenic with a single intraperitoneal injection of 200 mg/kg cyclophosphamide (CPA). Beginning 24 h later and for 4 consecutive days, the mice were given a daily intravenous injection of 100 μg/kg of an hG-CSF analog, met hG-CSF produced in our laboratory, clinically available hG-CSF (filgrastim, Amgen), or saline. On day 5, 6 h after the final dose, the animals were killed, blood samples were collected, and granulopoietic activity was determined by counting the number of white blood cells and polymorphonuclear neutrophils.

Pharmacokinetics

Plasma concentrations of a designed hG-CSF analog or wild-type hG-CSF (filgrastim, Amgen) were determined following administration in cynomolgus monkeys. Animals were given a single intravenous injection of 5 μg/kg or daily subcutaneous injections of 5 μg/kg for 28 d. In the intravenous study, blood samples were collected at 0 (predose), 5, 15, and 30 min and 1, 2, 4, 6, 8, 12, and 24 h postdosing. In the subcutaneous studies, blood samples were collected at 0 (predose), 1, 2, 4, 6, 8, 12, and 24 h postdosing on day 1 and day 28. All samples were immediately placed on wet ice and centrifuged at 28°C. The resultant plasma was then frozen and

stored (−70°C). Plasma concentrations were determined using an enzyme-linked immunosorbent assay (Quantikine human G-CSF ELISA, R&D Systems, Minneapolis, MN), performed per manufacturers instructions except that samples were diluted in PBS, 5% nonfat dry milk, and 0.05% Tween 20, and the incubation was extended to overnight at 4°C. Plasma concentrations of the designed hG-CSF analog and filgrastim were estimated from their corresponding standard curves. Pharmacokinetic parameters were calculated by noncompartmental analysis. The terminal slope (λz) was estimated by linear regression through the last time points of the log concentration versus time curves and used to calculate the terminal half-life ($t_{1/2}$). The area under the curve from time of dosing through the last time point (AUC_{0-z}) was calculated by the linear trapezoid method.

Acknowledgments

We thank Dr. Belinda Avalos (Ohio State University) for kindly supplying the Ba/F3 cell line transfected with the hG-CSF receptor. We also thank Dr. Steven Adams (American College of Laboratory Animal Medicine) and LAB Preclinical Research Institute Inc., (Quebec, Canada) for conducting the monkey studies.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Aritomi, M., Kunishima, N., Okamoto, T., Kuroki, R., Ota, Y., and Morikawa, K. 1999. Atomic structure of the GCSF-receptor complex showing a new cytokine-receptor recognition scheme. *Nature* **401:** 713–717.

Avalos, B.R., Hunter, M.G., Parker, J.M., Ceselski, S.K., Druker, B.J., Corey, S.J., and Mehta, V.B. 1995. Point mutations in the conserved box 1 region inactivate the human granulocyte colony-stimulating factor receptor for growth signal transduction and tyrosine phosphorylation of p75c-rel. *Blood* **85:** 3117–3126.

Bishop, B., Koay, D.C., Sartorelli, A.C., and Regan, L. 2001. Reengineering granulocyte colony-stimulating factor for enhanced stability. *J. Biol. Chem.* **276:** 33465–33470.

Bolon, D.N. and Mayo, S.L. 2001. Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci.* **98:** 14274–14279.

Chang, C.C., Chen, T.T., Cox, B.W., Dawes, G.N., Stemmer, W.P., Punnonen, J., and Patten, P.A. 1999. Evolution of a cytokine using DNA family shuffling. *Nat. Biotechnol.* **17:** 793–797.

Chen, K.Q. and Arnold, F.H. 1991. Enzyme engineering for nonaqueous solvents: Random mutagenesis to enhance activity of subtilisin E in polar organic media. *Biotechnology* **9:** 1073–1077.

Crameri, A., Whitehorn, E.A., Tate, E., and Stemmer, W.P. 1996. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* **14:** 315–319.

Dahiyat, B.I. and Mayo, S.L. 1996. Protein design automation. *Protein Sci.* **5:** 895–903.

———. 1997a. De novo protein design: Fully automated sequence selection. *Science* **278:** 82–87.

———. 1997b. Probing the role of packing specificity in protein design. *Proc. Natl. Acad. Sci.* **94:** 10172–10177.

Desjarlais, J.R. and Handel, T.M. 1995. De novo design of the hydrophobic cores of proteins. *Protein Sci.* **4:** 2006–2018.

Dunbrack, R.L. and Karplus, M. 1993. Backbone-dependent rotamer library for proteins—an application to side-chain prediction. *J. Mol. Biol.* **230:** 543–574.

Grossmann, M., Leitolf, H., Weintraub, B.D., and Szkudlinski, M.W. 1998. A rational design strategy for protein hormone superagonists. *Nat. Biotechnol.* **16:** 871–875.

Hattori, K., Shimizu, K., Takahashi, M., Tamura, M., Oheda, M., Ohsawa, N., and Ono, M. 1990. Quantitative in vivo assay of human granulocyte colony-stimulating factor using cyclophosphamide-induced neutropenic mice. *Blood* **75:** 1228–1233.

Heikoop, J.C., van den Boogaart, P., Mulders, J.W., and Grootenhuis, P.D. 1997. Structure-based design and protein engineering of intersubunit disulfide bonds in gonadotropins. *Nat. Biotechnol.* **15:** 658–662.

Hellinga, H.W. and Richards, F.M. 1994. Optimal sequence selection in proteins of known structure by simulated evolution. *Proc. Natl. Acad. Sci.* **91:** 5803–5807.

Herman, A.C., Boone, T.C., and Lu, H.S. 1996. Characterization, formulation, and stability of Neupogen (Filgrastim), a recombinant human granulocyte-colony stimulating factor. *Pharm. Biotechnol.* **9:** 303–328.

Hill, C.P., Osslund, T.D., and Eisenberg, D. 1993. The structure of granulocyte-colony-stimulating factor and its relationship to other growth factors. *Proc. Natl. Acad. Sci.* **90:** 5167–5171.

Home, P.D., Barriocanal, L., and Lindholm, A. 1999. Comparative pharmacokinetics and pharmacodynamics of the novel rapid-acting insulin analogue, insulin aspart, in healthy volunteers. *Eur. J. Clin. Pharmacol.* **55:** 199–203.

Horan, T., Wen, J., Narhi, L., Parker, V., Garcia, A., Arakawa, T., and Philo, J. 1996. Dimerization of the extracellular domain of granulocyte-colony stimulating factor receptor by ligand binding: A monovalent ligand induces 2:2 complexes. *Biochemistry* **35:** 4886–4896.

Howey, D.C., Bowsher, R.R., Brunelle, R.L., and Woodworth, J.R. 1994. [Lys(B28), Pro(B29)]-human insulin. A rapidly absorbed analogue of human insulin. *Diabetes* **43:** 396–402.

Ishikawa, M., Iijima, H., Satake-Ishikawa, R., Tsumura, H., Iwamatsu, A., Kadoya, T., Shimada, Y., Fukamachi, H., Kobayashi, K., Matsuki, S., et al. 1992. The substitution of cysteine 17 of recombinant human G-CSF with alanine greatly enhanced its stability. *Cell Struct. Funct.* **17:** 61–65.

Jiang, X., Farid, H., Pistor, E., and Farid, R.S. 2000. A new approach to the design of uniquely folded thermally stable proteins. *Protein Sci.* **9:** 403–416.

Kolvenbach, C.G., Narhi, L.O., Philo, J.S., Li, T., Zhang, M., and Arakawa, T. 1997. Granulocyte-colony stimulating factor maintains a thermally stable, compact, partially folded structure at pH2. *J. Pept. Res.* **50:** 310–318.

Kraemer-Pecore, C.M., Wollacott, A.M., and Desjarlais, J.R. 2001. Computational protein design. *Curr. Opin. Chem. Biol.* **5:** 690–695.

Kuga, T., Komatsu, Y., Yamasaki, M., Sekine, S., Miyaji, H., Nishi, T., Sato, M., Yokoo, Y., Asano, M., Okabe, M., et al. 1989. Mutagenesis of human granulocyte colony stimulating factor. *Biochem. Biophys. Res. Commun.* **159:** 103–111.

Lovejoy, B., Cascio, D., and Eisenberg, D. 1993. Crystal structure of canine and bovine granulocyte-colony stimulating factor (G-CSF). *J. Mol. Biol.* **234:** 640–653.

Lowman, H.B. and Wells, J.A. 1993. Affinity maturation of human growth hormone by monovalent phage display. *J. Mol. Biol.* **234:** 564–578.

Lu, H.S., Clogston, C.L., Narhi, L.O., Merewether, L.A., Pearl, W.R., and Boone, T.C. 1992. Folding and oxidation of recombinant human granulocyte colony stimulating factor produced in *Escherichia coli*. Characterization of the disulfide-reduced intermediates and cysteine—serine analogs. *J. Biol. Chem.* **267:** 8770–8777.

Malakauskas, S. and Mayo, S. 1998. Design, structure and stability of a hyperthermophilic protein variant. *Nat. Struct. Biol.* **5:** 470–475.

Marshall, S.A. and Mayo, S.L. 2001. Achieving stability and conformational specificity in designed proteins via binary patterning. *J. Mol. Biol.* **305:** 619–631.

Mayo, S.L., Olafson, B.D., and Goddard III, W.A. 1990. Dreiding: A generic forcefield for molecular simulations. *J. Phys. Chem.* **94:** 8897–8909.

Okabe, M., Asano, M., Kuga, T., Komatsu, Y., Yamasaki, M., Yokoo, Y., Itoh, S., Morimoto, M., and Oka, T. 1990. In vitro and in vivo hematopoietic effect of mutant human granulocyte colony-stimulating factor. *Blood* **75:** 1788–1793.

Pokala, N. and Handel, T.M. 2001. Review: Protein design—where we were, where we are, where we're going. *J. Struct. Biol.* **134:** 269–281.

Reidhaar-Olson, J.F., De Souza-Hart, J.A., and Selick, H.E. 1996. Identification of residues critical to the activity of human granulocyte colony-stimulating factor. *Biochemistry* **35:** 9034–9041.

Shimaoka, M., Shifman, J.M., Jing, H., Takagi, J., Mayo, S.L., and Springer, T.A. 2000. Computational design of an integrin I domain stabilized in the open high affinity conformation. *Nat. Struct. Biol.* **7:** 674–678.

Stemmer, W.P. 1994. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370:** 389–391.

Street, A.G. and Mayo, S.L. 1999. Computational protein design. *Structure Fold. Des.* **7:** R105–109.

Strop, P. and Mayo, S.L. 1999. Rubredoxin variant folds without iron. *J. Am. Chem. Soc.* **121:** 2341–2345.

Young, D.C., Zhan, H., Cheng, Q.L., Hou, J., and Matthews, D.J. 1997. Characterization of the receptor binding determinants of granulocyte colony stimulating factor. *Protein Sci.* **6:** 1228–1236.

Zhao, H., Giver, L., Shao, Z., Affholter, J.A., and Arnold, F.H. 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.* **16:** 258–261.

estimated between 20,000 and 40,000 metric tons (*10*, *12*). Assuming that pre-Incan technologies operated at efficiencies comparable to *huayras* and recognizing the initial presence of grades as rich as 25% Ag (*11*), our data imply that several thousand tons of silver were produced in pre-Incan times. Although major new archaeological discoveries in the Andes remain a distinct possibility, the likelihood seems equally probable that most of this silver was recycled and transported elsewhere in the Americas before conquest, or eventually exported overseas by the Spanish.

References and Notes

1. H. Lechtman, in *The Coming of the Age of Iron*, P. J. Wertime, J. D. Muhly, Eds. (Yale Univ., New Haven, CT, 1980), pp. 267–334.
2. R. L. Burger, R. B. Gordon, *Science* **282**, 1108 (1998).
3. E. P. Benson, Ed., *Pre-Columbian Metallurgy of South America* (Dumbarton Oaks, Washington, DC, 1979).
4. I. Shimada, S. Epstein, A. K. Craig, *Science* **216**, 952 (1982).
5. H. Lechtman, in *Tiwanaku and Its Hinterland: Archaeology and Paleoecology of an Andean Civilization Vol. 2*, A. L. Kolata, Ed. (Smithsonian Institution, Washington, DC, 2002), pp. 404–434.
6. K. O. Bruhns, *Ancient South America* (Cambridge Univ., Cambridge, 1994).
7. W. E. Rudolph, *Am. Geogr. Soc.* **26**, 529 (1936).
8. M. Vuille, *Int. J. Climatol.* **19**, 1579 (1999).
9. Materials and methods are available as supporting material on *Science* Online.
10. P. J. Bartos, *Econ. Geol.* **95**, 645 (2000).
11. W. E. Wilson, A. Petrov, *Miner. Rec.* **30**, 9 (1999).
12. P. J. Bakewell, *Miners of the Red Mountain: Indian Labor in Potosí 1545–1650* (Univ. of New Mexico, Albuquerque, NM, 1984).
13. R. Peele, *Sch. Mines Q.* **15**, 8 (1893).
14. I. Renberg, I. M. Wik-Persson, O. Emteryd, *Nature* **368**, 323 (1994).
15. M. L. Brännvall, R. Bindler, O. Emteryd, I. Renberg, *J. Paleolimnol.* **25**, 421 (2001).
16. C. Cobeli, *Environ. Sci. Technol.* **33**, 2953 (1999).
17. I. Rivera-Duarte, A. R. Flegal, *Geochim. Cosmochim. Acta* **58**, 3307 (1994).
18. B. B. Wolfe, *Palaeogeogr. Palaeoclim. Palaeoecol.* **176**, 177 (2001).
19. A. S. Ek, I. Renberg, *J. Paleolimnol.* **26**, 89 (2001).
20. A. F. Bandelier, *The Islands of Titicaca and Koati* (Hispanic Society of America, New York, 1910).
21. P. R. Williams, *World Archaeol.* **33**, 361 (2002).
22. A. L. Kolata, *The Tiwanaku: Portrait of an Andean Civilization* (Blackwell, Oxford, UK, 1993).
23. M. B. Abbott, M. W. Binford, M. Brenner, K. R. Kelts, *Quat. Res.* **47**, 169 (1997).
24. B. S. Bauer, C. Stanish, *Ritual and Pilgrimage in the Ancient Andes: The Islands of the Sun and the Moon* (Univ. of Texas, Austin, TX, 2001).
25. L. G. Thompson, E. Mosley-Thompson, J. F. Bolzan, B. R. Koci, *Science* **229**, 971 (1985).
26. A. K. Craig, in *Precious Metals, Coinage, and the Changes in Monetary Structures in Latin America, Europe, and Asia*, E. van Cauwenberghe, Ed. (Leuven Univ. Press, Leuven, Netherlands, 1989), pp. 159–183.
27. G. E. Ericksen, R. G. Luedke, R. L. Smith, R. P. Koeppen, F. Urquidi, *Episodes* **13**, 5 (1990).

28. Supported by the U.S. NSF-ESH (M.B.A), Natural Sciences and Engineering Research Council of Canada (A.P.W.) and Geological Society of America. We are grateful to H. Lechtman, M. Bermann, C. Cooke, and journal reviewers for their insightful comments; G. Seltzer for assistance in the field; and S. Root for support in the laboratory.

Supporting Online Material
www.sciencemag.org/cgi/content/full/301/5641/1893/DC1
Materials and Methods
Fig. S1
Tables S1 and S2

9 June 2003; accepted 18 August 2003

Inactivation of TNF Signaling by Rationally Designed Dominant-Negative TNF Variants

Paul M. Steed,* Malú G. Tansey,*† Jonathan Zalevsky,* Eugene A. Zhukovsky, John R. Desjarlais, David E. Szymkowski, Christina Abbott, David Carmichael, Cheryl Chan, Lisa Cherry, Peter Cheung, Arthur J. Chirino, Hyo H. Chung, Stephen K. Doberstein, Araz Eivazi, Anton V. Filikov, Sarah X. Gao, René S. Hubert, Marian Hwang, Linus Hyun, Sandhya Kashi, Alice Kim, Esther Kim, James Kung, Sabrina P. Martinez,† Umesh S. Muchhal, Duc-Hanh T. Nguyen, Christopher O'Brien, Donald O'Keefe, Karen Singer, Omid Vafa, Jost Vielmetter, Sean C. Yoder, Bassil I. Dahiyat‡

Tumor necrosis factor (TNF) is a key regulator of inflammatory responses and has been implicated in many pathological conditions. We used structure-based design to engineer variant TNF proteins that rapidly form heterotrimers with native TNF to give complexes that neither bind to nor stimulate signaling through TNF receptors. Thus, TNF is inactivated by sequestration. Dominant-negative TNFs represent a possible approach to anti-inflammatory biotherapeutics, and experiments in animal models show that the strategy can attenuate TNF-mediated pathology. Similar rational design could be used to engineer inhibitors of additional TNF superfamily cytokines as well as other multimeric ligands.

TNF is a proinflammatory cytokine that can complex two TNF receptors, TNFR1 (p55) and TNFR2 (p75), to activate signaling cascades controlling apoptosis, inflammation, cell proliferation, and the immune response (*1*–*5*). The 26-kD type II transmembrane TNF precursor protein, expressed on many cell types, is proteolytically converted into a soluble 52-kD homotrimer (*6*). An elevated serum level of TNF is associated with the pathophysiology of rheumatoid arthritis (RA), inflammatory bowel disease, and ankylosing spondylitis (*1*, *7*, *8*), and molecules that inhibit TNF signaling have demonstrated clinical efficacy in treating some of these diseases (*9*, *10*).

We have engineered dominant-negative TNF (DN-TNF) variants that inactivate the native homotrimer by a sequestration mechanism that blocks TNF bioactivity (fig. S1). Protein design automation (PDA), an in silico method that predicts protein variants with improved biological properties (*11*–*13*), was used to introduce single or double amino acid changes into TNF (Fig. 1A) to generate the desired biological profile while maintaining the overall structural integrity of the molecule. Specifically, our goal was to design homotrimeric TNF variants that (i) have decreased receptor binding, (ii) sequester native TNF homotrimers from TNF receptors by formation of inactive native:variant heterotrimers, (iii) abolish TNF signaling in relevant biological assays, and (iv) are easily expressed and purified in large quantities from bacteria. Variants were tested for TNF receptor activation in cell-based assays, and nonagonistic variants were then checked for their ability to antagonize native TNF in cell and animal models. Subsequently, we evaluated assembly state, receptor binding, and heterotrimer formation for several variants.

The computational design strategy used crystal structures of native and variant TNF trimers as templates for the simulations. Analysis of a homology model of the TNF-receptor complex revealed several distinct regions of the cytokine that make multiple direct contacts with its receptors (Fig. 1A), including interfaces rich in hydrophobic and electrostatic interactions. We ran simulations to select nonimmunogenic point mutations that would disrupt receptor interactions while preserving the structural integrity of the TNF variants and their ability to assemble into heterotrimers with native TNF (*14*). Many of the designed TNF variants displayed markedly reduced binding to TNFR1 and TNFR2, and several combinations of potent single mutations further decreased binding (Fig. 1B and fig. S2). As predicted by analysis of the TNF-TNFR structural complex, combinations of the most potent single mutations at different interaction domains (e.g., A145R and

Xencor, 111 West Lemon Avenue, Monrovia, CA 91016, USA.

*These authors contributed equally to this work.
†Present address: Department of Physiology, University of Texas Southwestern Medical Center at Dallas, Dallas, TX 75390, USA.
‡To whom correspondence should be addressed. E-mail: baz@xencor.com

I97T) were frequently additive or synergistic. Moreover, our data extend results of previous studies (*15*, *16*) in demonstrating that certain substitutions can alter the specificity of receptor interactions; for example, I97T and A145R/I97T show greater relative binding to TNFR2 than to TNFR1 (Fig. 1B and fig. S2).

Homotrimers of several designed TNF variants exhibited >10,000-fold reduction in their ability to activate two major signaling pathways downstream of TNF receptor activation. Specifically, single variants such as Y87H and A145R and the double variant A145R/Y87H were unable to bind to either TNFR1 or TNFR2 in cell-free assays (Fig. 1B and fig. S2) and failed to activate either caspase (Fig. 1C) or nuclear factor κB (NF-κB)–mediated luciferase expression (Fig. 1D) relative to native TNF. In contrast, those variants that still bound TNFRs also activated TNF signaling; in some cases (e.g., F144N) more potently than native TNF. Disruption of two receptor interfaces (e.g., A145R/Y87H or A145R/I97T) effectively destroyed the residual agonism detected with some single-point TNF variants. Furthermore, the importance of using multiple screening criteria to evaluate DN-TNF bioactivity was revealed by variants such as A145R/I97T, which had virtually no TNF-like activity in either cell-based assay yet displayed appreciable TNFR2 binding affinity.

We subsequently tested nonagonistic TNF variants for their ability to act as dominant-negative inhibitors by measuring their capacity to block native TNF activity in cell-based assays. We evaluated dose-dependent TNF antagonism (*14*) by mixing increasing concentrations of variants with native TNF (5 ng/ml) for 1.5 hours and measuring the caspase activity induced by these mixtures after addition to U937 cells (Fig. 2A). At concentrations as low as twofold that of native TNF (10 ng/ml), A145R/Y87H and

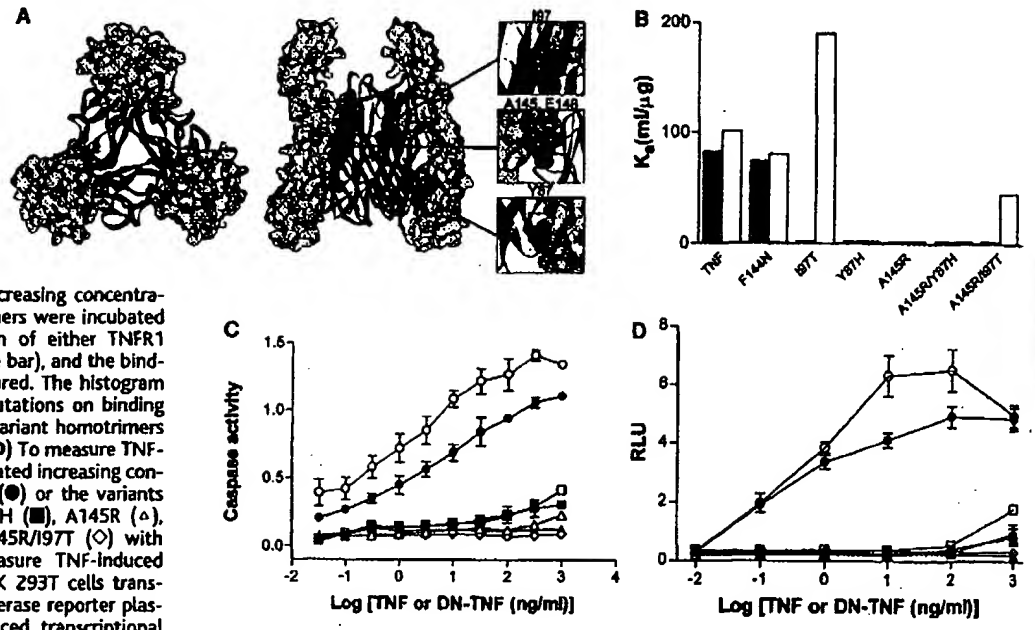


Fig. 1. DN-TNF variants have impaired TNF receptor binding and signaling. (A) Structural schematic of human TNF trimer–TNFR1 complex with major contacts between ligand and receptor highlighted by solid surfaces (green). Locations of representative mutated residues substituted in dominant-negative variants are shown in boxes. (B) Increasing concentrations of DN-TNF homotrimers were incubated with a fixed concentration of either TNFR1 (black bar) or TNFR2 (white bar), and the binding affinity (K_a) was measured. The histogram illustrates the effect of mutations on binding affinity between DN-TNF variant homotrimers and TNF receptors. (C and D) To measure TNF-induced signaling, we incubated increasing concentrations of native TNF (●) or the variants F144N (○), I97T (□), Y87H (■), A145R (△), A145R/Y87H (▲), and A145R/I97T (◇) with either U937 cells to measure TNF-induced caspase activity (C) or HEK 293T cells transfected with an NF-κB–luciferase reporter plasmid to measure TNF-induced transcriptional activation (D). DN-TNF variants, especially the double mutants, have reduced TNF receptor binding and signaling activity. RLU, relative luciferase units; Caspase activity, arbitrary units normalized to V_{max}.

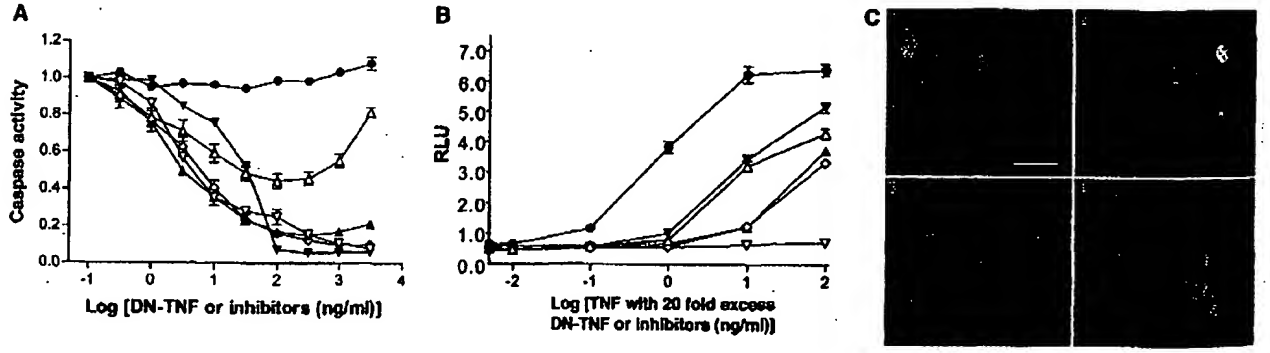


Fig. 2. DN-TNF variants inhibit TNF-mediated intracellular signaling. (A) Inhibition of caspase activation. Native TNF (5 ng/ml) was mixed with buffer (●) or with increasing concentrations of either TNF variants A145R (△), A145R/Y87H (▲), or A145R/I97T (◇), the soluble Fc-TNFR2 fusion etanercept (▽), or the TNF monoclonal antibody infliximab (▼). After 1.5 hours of incubation in exchange buffer (*14*), these mixtures were applied to U937 cells to stimulate caspase activity. (B) Inhibition of NFκB pathway activation. Native TNF (●) (25 μg/ml) was mixed with 20-fold excess (by mass) of A145R (△), A145R/Y87H (▲), A145R/I97T (◇), etanercept (▽), or infliximab (▼). These mixtures were serially diluted and applied to HEK 293T cells for 12 hours to induce NF-κB–luciferase reporter activity. (C) A145R/Y87H inhibits native TNF-induced nuclear translocation of the p65-RelA subunit of NF-κB. Immunofluorescence studies show subcellular localization of NF-κB in HeLa cells treated with buffer (panel 1), native TNF (10 ng/ml) (panel 2), A145R/Y87H (100 ng/ml) (panel 3), or the combination of native TNF and variant A145R/Y87H (panel 4). RLU, relative luciferase units; Caspase activity, arbitrary units normalized to V_{max}. Scale bar, 25 μm.

A145R/I97T attenuated TNF-induced caspase activity by 50%, and at 20-fold excess, activity was reduced to baseline. The in vitro potency (by mass) of these variants is comparable to that of a soluble Fc-TNFR2 fusion (etanercept) and more potent than that of an antibody to TNF (infliximab), two marketed anti-TNF therapies, supporting the potential utility of this mechanism. Similarly, at 20-fold excess over native TNF, single-point (A145R, I97T, Y87H) and particularly double-point (A145R/Y87H, A145R/I97T) variants decreased caspase activation (fig. S3) as well as TNF-induced transcriptional activation by NF-κB in human embryonic kidney (HEK) 293T cells (Fig. 2B). Consistent with these results, the TNF variant A145R/Y87H (at 10-fold excess over native TNF) blocked TNF-induced nuclear translocation of the NF-κB p65-RelA subunit in HeLa cells (Fig. 2C). Thus, a number of variants neutralized TNF-induced caspase and NF-κB–mediated transcriptional activity over a wide range of native TNF concentrations, including the clinically relevant range of 100 to 200 pg/ml found in the synovial fluid of RA patients (*17–19*).

To demonstrate that the mechanism of TNF inhibition requires the formation of heterotrimeric complexes with native TNF, we measured the relation between heterotrimer levels and inhibition of TNF-induced signaling (*14*). We generated heterotrimeric complexes by mixing a fixed amount of FLAG-tagged native TNF with increasing concentrations of His-tagged TNF variants. A part of this material was used in a sandwich enzyme-linked immunosorbent assay (ELISA) (Fig. 3A, open symbols) to detect the formation of His-FLAG heterotrimers, and the remainder was applied to U937 cells to detect TNF-mediated caspase activation (Fig. 3A, closed symbols). The extent of heterotrimer formation of A145R/Y87H or A145R/I97T with native TNF correlated with a decrease in caspase activation, demonstrating an inverse relation between signaling and heterotrimer formation. As expected, etanercept activity is independent of TNF monomer exchange (Fig. 3A, open circles) because etanercept binds to the TNF trimer. To directly visualize heterotrimer formation, we mixed FLAG-tagged native TNF with His-tagged DN-TNF and resolved the exchanged products using native polyacrylamide gel electrophoresis (PAGE) (Fig. 3B) (*20*). Electrophoresis of equimolar quantities of mixed DN-TNF and native TNF resolved the variant homotrimer, 1:2 and 2:1 native:variant heterotrimers, and native homotrimer in approximately the expected 1:3:3:1 ratio (Fig. 3B, lane 10:10). Western blot analyses (*14*) with antibodies against the epitope tags confirmed the composition of the intermediate species (fig. S4). Stochastic equilibrium modeling of native and variant TNF heterotrimer assembly predicts that 10-fold excess of variant homotrimer causes the loss of more than 99% of homotrimeric native TNF, primarily into 1:2 native:variant heterotrimers, and our results confirmed this (Fig. 3B, lane 10:100). Exchange reactions between native and variant TNF reached ~80% completion at 20 min, and essentially all the native homotrimer was depleted after 90 min (fig. S5). Finally, we confirmed that biological activity of variants requires exchange into heterotrimeric complexes with native TNF. Specifically, our most potent variants (e.g., A145R/Y87H) failed to block caspase activity induced by chemically cross-linked native TNF homotrimers (*14*), which are unable to dissociate to allow exchange with variant TNFs (fig. S6).

The most potent in vitro inhibitors were selected for testing in vivo, to further study the mechanism and to begin therapeutic lead candidate identification. We tested the bioactivity of variant homotrimer and native:variant heterotrimers in the D-galactosamine (GalN)-sensitized mouse model, which demonstrated that DN-TNF homotrimers, and heterotrimers with native TNF, are devoid of agonist activity and efficiently exchange with endogenous TNF in vivo. GalN is a known specific hepato-toxin that can increase the sensitivity of mice to human TNF by 1000-fold (*21*, *22*). Native human TNF (30 μg/kg) induced severe hepatocellular apoptosis and lethality, consistent with previous reports

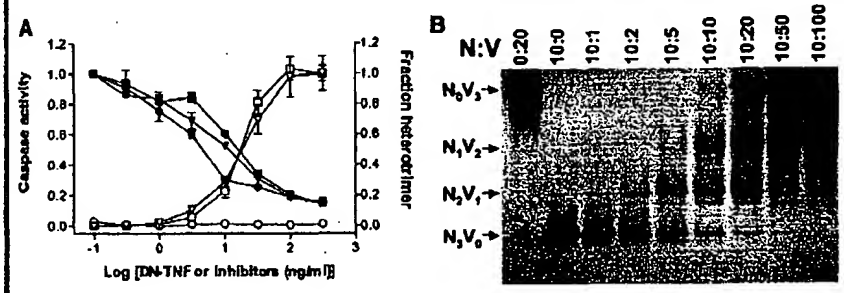


Fig. 3. DN-TNF variants inhibit signaling by sequestering native TNF into inactive heterotrimers. (A) Inverse correlation between heterotrimer formation and caspase activity. Native TNF was mixed in exchange buffer (*14*) with A145R/Y87H (▽, ▼), A145R/I97T (□, ■), or etanercept (○, ●), as described for Fig. 2. A part of each mixture was analyzed by a sandwich ELISA to detect native:variant TNF (open symbols), and the remainder was used to stimulate caspase activity in U937 cells (closed symbols). Caspase activity, arbitrary units normalized to V_{max}. (B) Native gel analysis of heterotrimer formation with various ratios of native (N) and DN-TNF (V). FLAG-tagged native TNF was incubated alone (N_3:V_0, lane 10:0) or with increasing concentrations of His-tagged variant A145R/Y87H (lanes 10:1 to 10:100) before native gel electrophoresis to determine heterotrimer formation. The differences in isoelectric point conferred by the epitope tags allowed for resolution of all possible trimer species (N_3:V_0, N_2:V_1, N_1:V_2, and N_0:V_3). Increasing concentrations of DN-TNF variant caused the redistribution of native TNF into both heterotrimers, and at 10-fold excess all detectable native TNF was consumed.

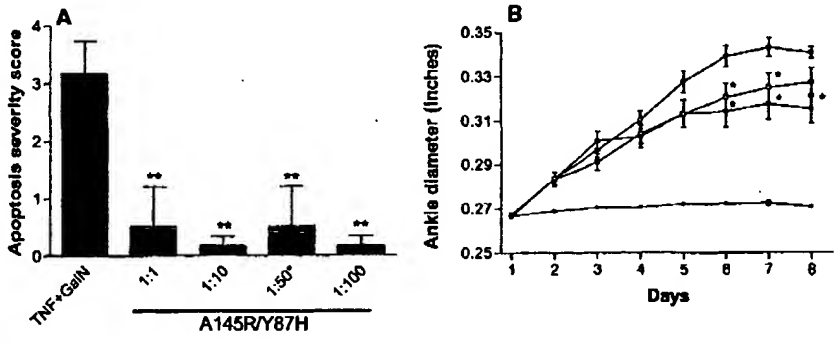


Fig. 4. DN-TNF variants exhibit efficacy in vivo. (A) Effect of heterotrimers of various ratios in the galactosamine-sensitized mouse model of human TNF-induced endotoxemia. Native human TNF was dosed at 30 μg/kg and A145R/Y87H was dosed at the indicated ratios to a fixed native human TNF dose of 30 μg/kg except at the 1:50* ratio, where A145R/Y87H was dosed in 50-fold excess of native human TNF (75 μg/kg). Livers were harvested and samples were blinded and scored for apoptotic damage on a scale of 0 to 4 as described (*14*). **$P < 0.05$. (B) Efficacy of A145R/Y87H in the rat 7-day established CIA model. A145R/Y87H was modified to introduce a PEG moiety at residue 31 of a non–epitope-tagged molecule as described (*14*). One group of four animals was nonarthritic (●); the remaining animals were collagen treated and, after the onset of symptoms, they were randomized into groups of eight. Animals were treated with vehicle (○), variant at 10 mg/kg twice daily dosing (□), or variant at 2 mg/kg subcutaneously with an intravenous loading dose of 2 mg/kg on the first treatment day (■). Measurements of ankle diameter were made daily by caliper. *$P < 0.05$.

(*23*); in contrast, A145R/Y87H dosed as high as 30 mg per kilogram of mouse body weight (along with native TNF at 30 μg/kg) resulted in no mortality or hepatotoxicity (Fig. 4A) (*24*). Similarly, lethal doses of native human TNF (30 μg/kg) mixed before injection with varying ratios of A145R/Y87H produced no TNF-induced damage. This protection was observed at native: variant ratios as low as 1:1 and with a superlethal dose of TNF (Fig. 4A). Further, sandwich ELISA analyses of serum samples indicated that a substantial portion (30 ng/ml) of administered A145R (3 mg/kg) was in heterotrimers with the endogenous mouse TNF at 1 hour.

A145R/Y87H was next assessed in a model of chronic disease as an initial test of the DN-TNF antagonism mechanism in a disease-relevant setting. We selected the rat 7-day established collagen-induced arthritis (CIA) model because it simulates chronic autoimmune joint disease and can be treated by TNF blockade (*25*). When dosed after the onset of symptoms, only interventions with rapid onset of action would be able to affect disease progression in this model, thus requiring rapid exchange in vivo of TNF variants with endogenous TNF. To ensure that there were no confounding in vivo effects of using affinity-tagged variants, we produced A145R/Y87H that lacked such tags. Further, to decrease in vivo clearance, we added one polyethylene glycol (PEG; ~5 kD/molecule) to each monomeric subunit of A145R/Y87H. This modification had no effect on the dominant-negative properties of the molecule in vitro (fig. S7). A145R/Y87H reduced joint swelling in the CIA model when dosed once daily at 2.0 mg/kg subcutaneously with a loading dose of 2.0 mg/kg and twice daily at 10 mg/kg intravenously (Fig. 4B). These results demonstrate the potential of DN-TNFs to inhibit TNF-mediated inflammation and verify that exchange occurs rapidly enough to affect progression of acute symptoms when dosed therapeutically.

Given their high-yield bacterial production, theoretical low immunogenicity, and unique mechanism of action, DN-TNFs show potential as a new class of anti-inflammatory therapy, particularly because existing methodologies (i.e., PEG modification) can be used to further enhance their pharmacokinetic properties (*26*, *27*). Further, we propose that this dominant-negative approach should be tested for its potential to create inhibitors of other multimeric extracellular signaling molecules, in particular other members of the TNF superfamily (e.g., RANKL, CD40L, and BAFF) that have been implicated in human pathophysiology (*28*, *29*).

References and Notes

1. B. B. Aggarwal, A. Samanta, M. Feldmann, in *Cytokine Reference*, J. J. Oppenheim, M. Feldmann, Eds. (Academic Press, London, 2000).
2. G. Chen, D. V. Goeddel, *Science* 296, 1634 (2002).
3. D. J. MacEwan, *Cell Signal.* 14, 477 (2002).
4. M. P. Boldin, T. M. Goncharov, Y. V. Goltsev, D. Wallach, *Cell* 85, 803 (1996).
5. G. M. Cohen, *Biochem. J.* 326, 1 (1997).
6. S. R. Ruuls *et al.*, *Immunity* 15, 533 (2001).
7. M. Feldmann, R. N. Maini, *Annu. Rev. Immunol.* 19, 163 (2001).
8. B. Beutler, *Immunity* 15, 5 (2001).
9. M. Feldmann, *Nature Rev. Immunol.* 2, 364 (2002).
10. R. Goldbach-Mansky, P. E. Lipsky, *Annu. Rev. Med.* 54, 197 (2003).
11. R. J. Hayes *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 99, 15926 (2002).
12. A. V. Filikov *et al.*, *Protein Sci.* 11, 1452 (2002).
13. P. Luo *et al.*, *Protein Sci.* 11, 1218 (2002).
14. Materials and methods are available as supporting material on Science Online.
15. J. Yamagishi *et al.*, *Protein Eng.* 3, 713 (1990).
16. X. M. Zhang, I. Weber, M. J. Chen, *J. Biol. Chem.* 267, 24069 (1992).
17. G. Steiner *et al.*, *Rheumatology* 38, 202 (1999).
18. T. Horiuchi *et al.*, *Endocr. J.* 46, 643 (1999).
19. A. K. Ulfgren *et al.*, *Arthritis Rheum.* 43, 2391 (2000).
20. P. Ameloot, W. Declercq, W. Fiers, P. Vandenabeele, P. Brouckaert, *J. Biol. Chem.* 276, 27098 (2001).
21. I. Hishinuma *et al.*, *Hepatology* 12, 1187 (1990).
22. S. Sakaguchi, S. Furusawa, K. Yokota, M. Takayanagi, Y. Takayanagi, *Int. J. Immunopharmacol.* 22, 935 (2000).
23. P. Brouckaert, C. Libert, B. Everaerdt, W. Fiers, *Lymphokine Cytokine Res.* 11, 193 (1992).
24. P. M. Stead *et al.*, data not shown.
25. A. M. Bendele *et al.*, *Arthritis Rheum.* 43, 2648 (2000).
26. K. Sreekrishna *et al.*, *Biochemistry* 28, 4117 (1989).
27. Y. P. Li *et al.*, *Biol. Pharm. Bull.* 24, 666 (2001).
28. C. F. Ware, *J. Exp. Med.* 192, F35-8 (2000).
29. R. M. Locksley, N. Killeen, M. J. Lenardo, *Cell* 104, 487 (2001).

30. We thank M. Ary for technical assistance with the manuscript.

Supporting Online Material
www.sciencemag.org/cgi/content/full/301/5641/1895/DC1
Materials and Methods
References
Figs. S1 to S7

9 December 2002; accepted 14 August 2003

The Dog Genome: Survey Sequencing and Comparative Analysis

Ewen F. Kirkness,[1] Vineet Bafna,[2*] Aaron L. Halpern,[2*] Samuel Levy,[2*] Karin Remington,[2*] Douglas B. Rusch,[2*] Arthur L. Delcher,[1] Mihai Pop,[1] Wei Wang,[1] Claire M. Fraser,[1] J. Craig Venter[2]

A survey of the dog genome sequence (6.22 million sequence reads; 1.5× coverage) demonstrates the power of sample sequencing for comparative analysis of mammalian genomes and the generation of species-specific resources. More than 650 million base pairs (>25%) of dog sequence align uniquely to the human genome, including fragments of putative orthologs for 18,473 of 24,567 annotated human genes. Mutation rates, conserved synteny, repeat content, and phylogeny can be compared among human, mouse, and dog. A variety of polymorphic elements are identified that will be valuable for mapping the genetic basis of diseases and traits in the dog.

Our understanding of how the human genome functions in health and disease will benefit from comparison of its structure with the genomes of other species (*1*, *2*). The domestic dog is a particularly good example, where an unusual population structure offers unique opportunities for understanding the genetic basis of morphology, behaviors, and disease susceptibility (*3*, *4*). The physical and behavioral characteristics of ~300 dog "breeds" are maintained by restricting gene flow between breeds. Many modern breeds are derived from few founders and have been inbred for desired characteristics. This has led to a species with enormous phenotypic diversity, but with significant homogenization of the gene pool within breeds. Many of the ~360 known genetic disorders in dogs resemble human conditions, and their causes may be more tractable in large dog pedigrees than in small, outbred human families (*4*, *5*). The combination of genetic homogeneity and phenotypic diversity also provides an opportunity to understand the genetic basis of many complex developmental processes in mammals (*6*).

Because of the costs of sequencing mammalian genomes to completion, these projects have been restricted to a few species that are considered to be of greatest value to biomedical research. The decision as to whether future projects should aim for complete sequence coverage of a few more genomes, or whether the existing "reference genomes" can be exploited to characterize a wider variety of genomes that are sequenced to a lower level of coverage, must be made. Here,

[1]The Institute for Genomic Research, Rockville, MD 20850, USA. [2]The Center for Advancement of Genomics, Rockville, MD 20850, USA.

*These authors contributed equally to this work.

PROTEINS TO ORDER

Program determines amino acid sequence from three-dimensional structure data

A computer program developed by researchers at California Institute of Technology can identify amino acid sequences capable of folding into the three-dimensional structure of a specific protein. The work could lead to a deeper understanding of the way protein sequence specifies structure and could also provide a new method for creating modified proteins with enhanced properties.

Chemistry graduate student Bassil I. Dahiyat and biology professor and Howard Hughes Medical Institute investigator Stephen L. Mayo used their computational design algorithm to design a synthetic protein that mimics the fold of a zinc finger, a naturally occurring DNA-binding element found in some transcription factors [*Science,* 278, 82 (1997)]. The natural zinc finger has only 28 residues, but it contains the three main secondary structures that make up all proteins (β-sheet, α-helix, and turn elements), plus a zinc ion. The Caltech researchers' synthetic zinc finger mimics the secondary structures but does not include a metal ion.

"Using an algorithm to solve the inverse folding problem—that is, to find sequences that will adopt a given fold—Dahiyat and Mayo have taken protein engineering to a new high," comments George D. Rose, professor of biophysics and biophysical chemistry at Johns Hopkins University.

"Creation of the de novo zincless finger involved authentic engineering: design, implementation, and confirmation that the product satisfied its design specs," Rose points out. "The successful execution of all three steps in this process has been a long time in coming." Rose and coworkers have developed one of the most successful algorithms for solving the opposite problem—determining the 3-D structure of a protein from its amino acid sequence.

Biochemistry professor William F. DeGrado of the University of Pennsylvania School of Medicine, who specializes in protein structure, folding, and design, writes in the same issue of *Science* that the zinc-finger mimic "is the smallest protein known to be capable of folding into a unique structure without the thermodynamic assistance of disulfides, metal ions, or other subunits. This important accomplishment illustrates the impressive ability of Dahiyat and Mayo's program to design highly optimized sequences."

The "spectacular success" achieved by Dahiyat and Mayo, DeGrado notes, suggests that "the rules of protein folding and computational methods for de novo design may now be sufficiently defined to allow the engineering of a variety of proteins."

DeGrado tells C&EN that the work also "brings us closer to the design of entirely nonnatural biomimetic polymers with predetermined 3-D structures. There are many components to Mayo's potential energy function, but the most important of them are based on atom-level interactions that are not specific to proteins. Hence, the program could be generalized to aid in the construction of a variety of polymers that fold in predictable geometries."

A protein with 28 amino acids can have 1.9×10^{27} different possible sequences. The task of the program devised by Dahiyat and Mayo was to select the optimal sequence from among these possibilities. The algorithm's underlying strategy is to repeatedly find and eliminate sequences that are a bad fit to the target structure so it can rapidly converge on a solution. The program assesses interactions between each side chain and the protein backbone, as well as interactions between side chains, and it uses a mathematical proposition called the dead-end elimination theorem to make its calculations computationally manageable.

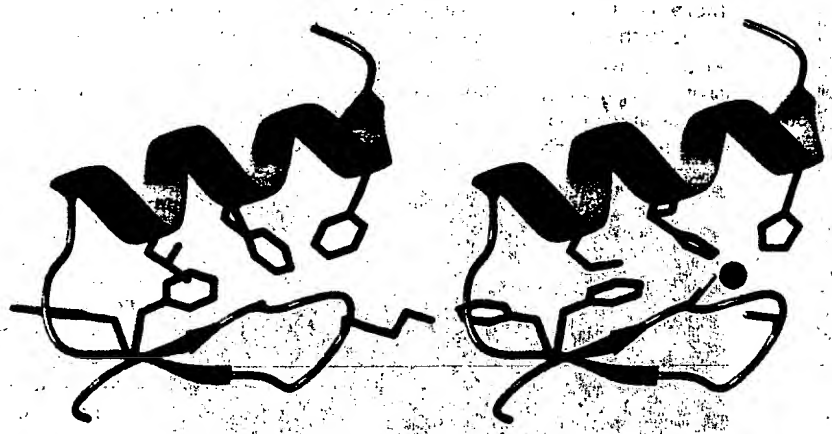
The sequence selected by the program was found experimentally to adopt a folded structure that closely approximates that of a zinc finger, even though the mimic's amino acid sequence bears little resemblance to that of a zinc finger, or that of any other known protein, for that matter. Six of the mimic's amino acid residues are identical to those of the natural protein, another five are similar, and the other 17 are completely different.

The zinc-finger mimic does not reproduce the natural protein's DNA-binding function. "Our goal was not to retain the function but just to retain the structure of the fold," says Mayo. "We are attempting, however, to design zinc-finger variants that do, in fact, still bind to DNA and perhaps change the DNA-binding specificity of the molecule, which would be very exciting."

The program's predictive abilities have so far been demonstrated only for the zinc finger. But the researchers believe the algorithm will prove capable of calculating sequences for a range of other proteins as well. Dahiyat has now moved to Xencor, Pasadena, Calif., a company founded to commercialize the new technology.

The protein-folding problem—how structure is derived from sequence—is among the most significant unsolved issues in chemistry and biology. "What we've been able to do is obviously not to solve the protein-folding problem," says Mayo, "but to make a large inroad into solving the inverse folding problem—which is, given some target structure, how do you design a sequence that will fold to that structure?" This knowledge could lead to a better understanding of the underlying physical chemistry of protein folding.

In addition, the technology should be useful for protein design and modification.



Structure of zinc-finger mimic synthesized by Dahiyat and Mayo (left) closely matches that of natural zinc finger protein. Purple sphere is zinc ion.

In preliminary work, Mayo and biology graduate student Sandra M. Malakauskas have been able to computationally design a mutant of the 56-residue Protein G that transforms it so that the protein remains fully folded even at 100 °C. "This has potential implications for developing thermostable enzymes for use in industrial processing," says Mayo.

Stu Borman

Cost of rail problems continues to mount

Huge losses and thousands of idled workers have raised chemicals and plastics manufacturers' complaints to Union Pacific (UP) railroad to a deafening roar.

Last year's merger of Southern Pacific with UP has led to service disruptions throughout the western U.S. and the Gulf Coast. With much chemical traffic having no alternative to rail, UP's problems are having a devastating effect on business, especially in the chemical industry-rich Gulf Coast region (C&EN, Sept. 1, page 17).

Rail cars stranded in Houston

The Chemical Manufacturers Association (CMA) now estimates losses at more than $100 million and growing at $36 million monthly for chemical producers in UP service areas.

"The railroads can make or break the plastics and chemical industry," says Society of the Plastics Industry (SPI) President Larry L. Thomas. "And for that reason SPI and CMA have no alternative but to take aggressive action to seek immediate relief."

In a joint statement released Oct. 1, CMA and SPI state: "We do not believe that UP alone can solve the problems it has created. We continue to urge the U.S. Surface Transportation Board [STB] to intervene in this growing transportation crisis."

At CMA's board of directors meeting in Williamsburg, Va., last month, UP Chairman and Chief Executive Officer Richard K. Davidson explained the company's plan for alleviating the situation to 50 chemical industry executives. Additional details were included in UP's quarterly merger report filed with STB, a division of the Department of Transportation.

Actions highlighted in the UP plan include temporary diversions of some rail traffic to other railroads and rerouting trains around congested terminals, including Houston. UP estimates that Gulf Coast service should return to normal in 60 to 90 days under this plan.

But CMA and SPI are not convinced. Both organizations have rejected UP's latest plan to relieve the congestion and are talking to other railroads, including Kansas City Southern and Burlington Northern Santa Fe, about increasing their operations in the Gulf Coast.

Several parties have called for STB to issue an emergency services order, forcing UP to temporarily turn over parts of its operations to the STB or other carriers.

CMA and SPI say they will continue to pressure UP and STB. CMA is surveying its full membership to compile a report on company losses due to lost or slowed production, alternate shipping methods, and idled equipment. The initial $36 million estimate in monthly losses reflects the experience of only 17 of CMA's 191 member companies, so that figure is expected to rise.

SPI is focusing on downstream plastics processors to gauge the losses for businesses that are experiencing delayed shipments from their chemical suppliers.

The congestion problems are only part of more pervasive problems at UP. After a series of fatal accidents this summer, the Federal Railroad Administration (FRA) began a safety review resulting in strong recommendations for changes in UP procedures. In recent weeks, UP appointed a new vice president for safety and risk management at FRA's suggestion and dismissed its vice president of operations.

Discussions of UP's problems continue. On Oct. 3, the Railroad Commission of Texas, the state that produces the largest volume of chemicals, began a series of hearings on the rail situation in Houston. Additional sessions are scheduled for Fort Worth, San Antonio, and El Paso in coming weeks, and chemical shippers are expected to be major participants.

Paige Morse

Climate change debate heating up

With December negotiations on a global climate change treaty in Kyoto, Japan, just two months away, more countries are revealing their goals for reducing greenhouse gas emissions, and claims and counterclaims on what it would cost to achieve those goals are fogging the air.

At the Kyoto meeting, Japan is planning to propose an overall reduction of 6.5% in emissions below 1990 levels by 2010. That's far less than the European Union wants. The EU announced last spring that it backs a 15% rollback in emissions by industrialized countries by 2010. The U.S.'s plan is likely to disappoint both. It is now considering three options: stabilization of emissions at 1990 levels by 2010, by 2020, or by 2030.

The U.S.'s first option is feasible, and achieving it won't harm the economy, concludes the Department of Energy (DOE) in a just-released, peer-reviewed study. If no strong efforts are made to cut back on the use of fossil fuels, the DOE's Energy Information Administration projects that U.S. carbon dioxide emissions (measured as carbon) will rise to 1.73 billion metric tons in 2010, up from the 1990 level of 1.34 billion metric tons.

The study, conducted by five DOE labs, says the U.S. could minimize the costs of stabilizing greenhouse gases by investing in technologies such as natural-gas-fueled turbines for generating electricity, vehicles that are more fuel efficient, and energy-saving appliances and building materials.

"This analysis shows that what's good for the environment also can be good for the economy," says Energy Secretary Federico F. Peña.

Specifically, the DOE study looks at energy technologies and a domestic system of carbon dioxide emissions trading (similar to the sulfur dioxide trading system currently in place to minimize acid rain) that could be used to stabilize carbon dioxide emissions. The costs of the policies examined range from $50 billion to $90 billion per year, and the energy cost savings range from $70 billion to $90 billion annually. Consequently, the net economic cost would be near zero.

"If you design a climate policy to allow time and flexibility, and if you adopt a technology development program, then we can use the ingenuity of our scientists and engineers to achieve more than people thought," says Joseph